

DR	InterPro:	IPR001798; Kelch.
DR	Fam:	PF00651; RTB; 1.
DR	Fam:	PF01344; Kelch; 6.
DR	PRINTS:	PR00501; KELCHREPEAT.
DR	SMART:	SM0225; RTB; 1.
DR	PROSITE:	PS0097; RTB; 1.
KW	Cyscoletion:	Artin binding; Repeat.
FT	DOMAIN	45 90 SER RICH.
FT	DOMAIN	215 282 RTB.
FT	REPEAT	463 509 KELCH 1.
FT	REPEAT	510 556 KELCH 2.
FT	REPEAT	558 603 KELCH 3.
FT	REPEAT	604 650 KELCH 4.
FT	REPEAT	652 703 KELCH 5.
FT	REPEAT	704 750 KELCH 6.
SQ	SEQUENCE	751 AA: BZP32 MW: 668461907AVQ4F3 CRG64.

Query Match		5, 4%	Score 87.5;	DB 1; Length 751;
Best Local Similarity		22.0%;	Prod. No. 4;	
Matches	70;	Conserved 100	42;	Mismatches 9/; Indels 109; Gaps 10.

QY	10	ASSTGVKQVAFIKGHIAVFAAKKAQTPEELSVETFEVTHENLQGEQMAYE	69
DB	156	ASSI--- - QATGGQH RLISINSLTDSID	194
QY	70	LLENFGDTLMGSLERKYLAWLYAYKR	126
DB	194	AVRIAE QSFRKMNYLK QQQITIVLVNRKTPAHRLVS SVSYFAAMPTSD	248
QY	127	MEDEEDKYKE IFFWAIGSWRELFPQPIKH	167
DB	246	PAKDEEFKMSHHPNAILMLVFATCTGLKKETLERAAATLGLTVVVVTHFL	408
QY	163	AK --- MYRAVSSRGCDVMGSKDTHAWLRDMRMHSVAMP	212
DB	309	KLHLHSNCLTRAFALAGCTELMK	354
QY	213	PRGEGLTFSATCIRKACVLDSSRVSNINSSEGE ITH	268
DB	355	ILTFAP ELRLKLSNTNV	400
QY	264	LPPELLDLDF Z73	
DB	401	LATLELLQLLALE 418	

RESULT 6

ID	YP24-CABEL	STANDARD;	PRT:	446 AA.
AC	G09341;			
DI	01-NOV-1997	(Rel. 35, Created)		
DPT	01-NOV-1997	(Rel. 35, Last sequence update)		
DEF	01-NOV-1997	(Rel. 35, Last annotation update)		
DE	Hypothetical 40.0 kDa protein W0283.5 in chromosome III			
GEN	W0283.5			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;			
ACC	Rhabditidae; Pelodididae; Caenorhabditis			
ORX	NHL TaxID 6239;			
SRN				
REP	SEQUENCE FROM N.A.			
BCC	STRAIN BRISTOL N2;			
MA	Camell M.;			
REL	Submitted (JUL-1995) to the EMBL/Genbank/DDBJ databases.			
NC	1; SIMILARITY: TO C.ELEGANS W0283.5 AND TO THE TERMINAL OF T07A03.1			

this SwissProt entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL databank. The European Bioinformatics Institute. There are no restrictions on the use by non profit institutions as long as its content is in no way modified and this statement is not removed, displayed and for commercial entities requires a license agreement. http://www.ebi.ac.uk/seqdb/doc/

DB 669 KIPENFWAEKNY...LSATVDELNKK...VNIVKNPEETI- 406
 QY 214 KOPGLTPASCTLCHKAGVCSGVSNSNSSSPEQIFHYTNKFWSSQELMLDPELLIDP- 272
 DB 407 -----AEIKNKGILSDLELKKFDEVKVSLNKN- PRELTWDEA 443
 QY 274 ---ECHDILLLOEPLVGLPEPG 292
 DB 444 EMIVNCIQSMFMAPIVIGLRPIG 467
 RESULT 10
 CHD1 YEAST
 ID CHD1_YEAST STANDARD PRT 1468 AA
 AC P32657;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01 OCT 1993 (Rel. 27, Last sequence update)
 DT 01 MAR-2002 (Rel. 41, Last annotation update)
 DE CHD1 protein.
 GN CHD1 OR YER164W OR SYG1-ORF4.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 CX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mulligan J.T., Dietrich F.S., Hennessy K.M., Schi P., Komp C.,
 RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AH972;
 RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Allen E., Araujo R.,
 RA Aviles E., Herno A., Houtman T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew D., Liu D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oliver P., Oh C.,
 RA Patel F.X., Roberts D., Schi P., Schramm S., Shapiro T., Smith V.,
 RA Taylor P., Wei Y., Yellon M., Hotstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SOURCE/LOCUS: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SNF2/PAC54 HELICASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
 CC this SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.slb.ch/announcements/>
 CC or send an email to license@isb.slb.ch)
 CC
 DB EMBL: U18917; AB064691.1;
 DB PIR: S30818; S30818.
 DB SGD: S0000966; CHD1.
 DB InterPro: IPR000953; Chromo.
 DB InterPro: IPR001410; DEAD.
 DB InterPro: IPR001650; Helicase_C.
 DB InterPro: IPR000430; SNF2_N.
 DB Pfam: PF00485; chromo_2.
 DB Pfam: PF00271; helicase_C_1.
 DB Pfam: PF00176; SNF2_N_1.
 DB SMART: SM00298; CHROMO_2.
 DB SMART: SM00487; DEKDC_1.
 DB SMART: SM00490; HELIC_1.
 DB PROSITE: PS00598; CHROMO_1; 2.
 DB PROSITE: PS50014; CHROMO_2; 2.
 DB PROSITE: PS50090; MYH_3; UNKNOWN_1.
 KW DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
 FT DOMAIN 195 257 CHROMO 1.
 FT DOMAIN 285 350 CHROMO 2.
 FT NF_BIND 401 408 ATP (POTENTIAL).

ET SITE 514 516 DEAD BOX.
 SQ SEQUENCE 1468 AA; 168240 MW; 288007477ED96B536964;
 Query Match 5.1%; Score 84; DB 1; Length 1408;
 Best Local Similarity 20.6%; Pred. No. 18;
 Matches 74; Conservative 60; Mismatches 12%; Indels 90; Gaps 17;
 QY 19 KGVICRCHPHAVVCTAPKALDPRF PVSYPVQVVAETHNINCPDPAALPDLDPDQ 77
 DB 1072 KELLKELKHAI AYKAKKSTKAKNGKDNH LKQSPKPKKA VLNPKGV 119;
 QY 78 FFLSMIST PISP-KYLIAVLAVPRRAALTYTSEYTMNITVALYLAHMEHLELYK 140
 DB 1126 KSLNAESLSKVELAYLKNLINSNKIKKQKESLNNIK PVGNKSNWIKLEKELT 1184
 QY 146 YEFPWALGDSWRELPD FIKRDLDFWAKMNTVAVSR 174
 DB 1185 IGVKRYGV SWTQIRDPPLRTITRTETNFVNVPVAKFSASSSTETTPSEKPEKELT 1244
 QY 174 ---RCDVMS --- KQTHWAWLMDR PMHSCAMMS 204
 DB 1244 GKVTGALHGRKVDLGLSKGLMLKSLGSLGSKKLTGSKKPKPKPKPKSKMP 1304
 QY 264 YLRREI DETERGES LTPASCTLCHKAGVDSGQVSNSSSPE 246
 DB 1404 KESSEFANVDSKWKALDQTPAALINNIETTPSPHPEKSKVSKWET 1464
 QY 247 QETPHYINRHSQELMLDPELLIDPCTPDLDTQFQWGLPELDTAL FW 297
 DB 1362 SPSAH- PREY -----DSMEEPWTHPCATSELEKELRERKSLDREW 1404
 RESULT 11
 CRP_RABIT
 ID CRP_RABIT STANDARD PRT 225 AA.
 AC P02742;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE C-reactive protein precursor.
 GN CRX1 OR CRP.
 OS Erytholobus curvulus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-NEW ZEALAND WHITE; TISSUE-Liver;
 RC MEDLINE-87101075; PubMed-6026463;
 RA Wu S., Miller S.M., Samols D.;
 RT Cloning and characterization of the gene for rabbit C-reactive
 RT protein.*;
 RL Biochemistry 25:7834-7839(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-86168292; PubMed-3007466;
 RA Sym C., Gotschlich E.C., Liu L. Y.;
 RT Rabbit C-reactive protein. Biosynthesis and characterization of cDNA
 RT clones.*;
 RL J. Biol. Chem. 261:5473-5479(1986).
 RN [3]
 RP SEQUENCE OF 21-225.
 RC MEDLINE-84056861; PubMed-6754715;
 RA Wang C. M., Nguyen N.Y., Yamada K., Kobay F., Liu L. Y.;
 RT Primary structure of Rabbit C-reactive protein.*;
 RL J. Biol. Chem. 257:13610-13615(1982).
 CC -1- FUNCTION: CRP DISPLAYS SEVERAL FUNCTIONS ASSOCIATED WITH HOST
 CC PHAGOCYTOSIS, & COMPLEMENT FIXATION THROUGH ITS CALCIUM-DEPENDENT
 CC BINDING TO PHOSPHOTICHOINE.
 CC -1- SUBUNIT: HOMOPENTAMER. PENTAXIN (OR PENTAXIN) HAVE A 1035-1110
 CC ARRANGEMENT OF 5 NON-EQUIVALENT BINDING SUBUNITS.

CC -1 INDUCTION: THE CONCENTRATION OF THE IN PLASMA INCREASES GREATLY
 CC DURING ACUTE PHASE RESPONSE TO TISSUE INJURY OR INFLAMMATION.
 CC -2 MISCELLANEOUS: ASP-6 (AR 76, ARC-7), AND JLU-81 MAY BE INVOLVED
 CC IN THE CALCIUM-DEPENDENT BINDING OF PHOSPHOLIPOLIPIN, A PROTEIN
 CC THAT MAY BE IMPORTANT FOR THE BIOLOGICAL FUNCTION OF THIS PROTEIN.
 CC -3 SIMILARITY: BELONGS TO THE PENTAXIN FAMILY.
 CC
 CC THIS SWISS PRO entry is complete. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announcements>
 CC or send an email to license@isb.ch).

DR EMBL: M14538; AAA75403.1; 225
 DR EMBL: I47237; AAA75404.1; 225
 DR EMBL: M1497; AAA3106.1; 225
 DR PIR: A01202; GJRR.
 DR PIR: A25583; A25583.
 DR PIR: A25605; A25605.
 DR HSSP: P02741; ICBV.
 DR InterPro: IPR001759; Pentaxin.
 DR Pfam: PF00354; pentaxin_1.
 DR PRINTS: PF00895; PENTAXIN.
 DR ProDom: PD002153; Pentaxin; 1
 DR SMART: SM00159; PTX; 1.
 DR ProSITE: PS00289; PENTAXIN; 1
 KW Acute phase; Plasma; Calcium; Ventral; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 225
 FT DOMAIN 21 225
 FT DISULFID 55 116
 FT VARIAT 62 62
 FT VARIAT 89 89
 FT CONFLICT 10 10
 FT CONFLICT 46 46
 FT CONFLICT 50 50
 FT CONFLICT 76 76
 FT CONFLICT 84 88
 FT CONFLICT 91 91
 FT CONFLICT 93 101
 FT CONFLICT 104 108
 FT CONFLICT 167 167
 FT CONFLICT 177 177
 FT CONFLICT 193 193
 FT CONFLICT 225 225
 SQ SEQUENCE 225 AA; 2549 MW; 7025008336074962 CRC64.

Query Match 5.1%, Score 83, DB 1, Length 225,
 Best local Similarity 23.8%, Prod. No. 2 27, Indels 79, Gaps 15,
 Matches 64; Conservative 30; Mismatches 67.

QY 79 FLMSCLRSKYLKIAMVLAVEREASHLY SEY YITIN -----FFVALYLAN 125
 DB 8 FLVLSFNNMQD-----AGMHAFAVUKESUNSVSLNAQLKKPKAKFTVCLYPT 60
 QY 129 LMBHREYKYLIFWALGWSH LPL LPTKAK MVRKAVVLPDCLVWTK 142
 DB 61 LLSMTK--GYTFYAT--PGLNFTL--LWKLGYSFSSV--PRTIKVS 106
 QY 184 LQPTH--W-----AWLEIRPMHSGAMPYL-----RNEDDP--EP 213
 DB 107 LHVQTHPCASWSESLAGWVGRMAKSKFVYHGLGASLHQQWSPGSEF 166
 QY 214 LPLTEASLILHAGVDSGVGQVQV PGLIHILHSEVQLL 252
 DB 167 LKLSL-----VPLGNVAVVVA SPRLNTIYATGTSFNVL 204

RESULT 12
 CC100% ANAL STANDARD; ERT; 497 AA.
 ID CC100% ANAL

F49827;
 01 FEB 1995 (Rel. 31, Created)
 01 FEB 1995 (Rel. 31, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Cell division control protein 10.
 CD10.
 Candida albicans (Yeast).
 Eukaryota; Fungi; Ascomycota, Saccharomycotina, Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 NR91_TaxID=5476;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=94201183; PubMed=8152419;
 FA Didomenico B.J., Brown N.H., Lupisella J., Greene J.R., Yanko M.,
 Koltin Y.;
 "Homologs of the yeast neck filament associated genes: isolation and
 sequence analysis of Candida albicans CDC3 and CDC10.";
 Mol. Gen. Genet. 242:689-694(1994).
 -1- FUNCTION: PLAYS A ROLE IN THE CELL CYCLE. INVOLVED IN THE
 FORMATION OF THE RING OF FILAMENTS IN THE NECK REGION AT
 THE MOTHER-NUD JUNCTION DURING MITOSIS.
 -2- SUBCELLULAR LOCATION: PRESENT AT THE HUD NECK DURING CELL
 DIVISION (BY SIMILARITY).
 -3- DEVELOPMENTAL STAGE: EXPRESSED AT HIGHER LEVELS IN CELLS GROWING
 AS HYPHAE THAN IN THOSE GROWING AS BUDDING YEASTS.
 -4- SIMILARITY: BELONGS TO THE SEPTIN FAMILY.
 This SWISS-PRO entry is copyright, it is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announcements>
 or send an email to license@isb.ch).

EMBL: Z25870; CAA81090.1; 357
 InterPro: IPR000038; GTP_Cell_Div.
 Pfam: PF00735; GTP_CDC; 1.
 ProDom: PD002565; GTP_Cell_Div; 1.
 KW Cell division; Cell cycle; GTP-binding.
 FT NP_BIND 44 51 GTP (POTENTIAL).
 FT SEQUENCE 357 AA; 40654 MW; 7025008336074962 CRC64;
 Query Match 5.1%, Score 83, DB 1, Length 357,
 Best local Similarity 21.7%, Prod. No. 4.2,
 Matches 38; Conservative 35; Mismatches 70; Indels 32; Gaps 7;

QY 27 PHAVVVARPAQIPEPEELSVFKMVENTIILN-----LQDEEQAYRLENE 74
 DB 117 PLVYKREQSQYKRLKLAQKPKLADIKVHGLYFIPNGGKFLQIDVQAKKLSIA 176
 QY 75 QIQEFLSMDSCLRISDKYLIAWLAVERAAGLYSTVITNFFVALYLANDMEEDF--- 141
 DB 177 NVVPLIAKSDSLTDER-----SEPKK---LQSEFMKNINIIDYSELYEERQL 226
 QY 172 LKARYLIFWALGWSH LPL LPTKAK MVRKAVVLPDCLVWTK 183
 DB 61 LLSMTK--GYTFYAT--PGLNFTL--LWKLGYSFSSV--PRTIKVS 106
 QY 184 LQPTH--W-----AWLEIRPMHSGAMPYL-----RNEDDP--EP 213
 DB 107 LHVQTHPCASWSESLAGWVGRMAKSKFVYHGLGASLHQQWSPGSEF 166
 QY 214 LPLTEASLILHAGVDSGVGQVQV PGLIHILHSEVQLL 252
 DB 167 LKLSL-----VPLGNVAVVVA SPRLNTIYATGTSFNVL 204

RESULT 13
 CC100% ANAL STANDARD; PRI; 6669 AA.
 ID CC100% ANAL
 01-FEB-1991 (Rel. 17, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Nuclein.
 NFI.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NBI_TaxID=9606;
RP [1]
RX MEDLINE=95257391; PubMed=7739042;
RA Label S., Kolmerer H.;
RT "The complete primary structure of human nebulin and its correlation
to muscle structure";
RL J. Mol. Biol. 248:308-315(1995).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE.
RX MEDLINE=88284704; PubMed=4397062;
RA Zeviani M., Barrias H.T., Rizzuto K., Salvati G., Botton R.,
RA Bonilla E., Miranda A.F., Du J., Samitt C., Dickson G., Walsh F.S.,
RA Dinardo S., Francke U., Schon E.A.;
RT "Cloning and expression of human nebulin cDNAs and assignment of the
gene to chromosome 2q31-q42";
RL Genomics 2:249-256(1988).
RN [3]
RP STRUCTURE BY NMR OF 6610-6669.
RX MEDLINE=98179559; PubMed=9514727;
RA Politou A.S., Millevol S., Gautel M., Kolmerer H., Pastore A.;
RT "SH3 in muscles: solution structure of the SH3 domain from nebulin";
RL J. Mol. Biol. 276:189-202(1998).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
CC ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
CC OF STRIATED MUSCLE.
CC -1- DISEASE: Defects in NBR are a cause of the autosomal recessive
CC form of nemaline myopathy (NEM2).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN
CC -1- SIMILARITY: CONTAINS 178 NEBULIN REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/usage/>
CC or send an email to license@sib.ch).
CC
DR EMBL: X84957; CAA58788.1;
DR EMBL: M19668; AAA59916.1; A11_SHQ.
DR EMBL: M19669; AAA59917.1; ALT_SEQ.
DR PIR: A29979; A29979.
DR PIR: B29979; B29979.
DR IUB: LARK; 28-JAN-98.
DR IUB: INER; 24-DEC-97.
DR MIM: 161650;
DR MIM: 256030;
DR InterPro: IPR000900; Nebulin.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00880; Nebulin_repeat; 143.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00510; NEBULIN.
DR SMART: SM00227; NBR; 178.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 76 107 NEBULIN 1.
FT REPEAT 113 143 NEBULIN 2.
FT REPEAT 148 178 NEBULIN 3.
FT REPEAT 183 213 NEBULIN 4.
FT REPEAT 218 248 NEBULIN 5.
FT REPEAT 253 283 NEBULIN 6.
FT REPEAT 289 318 NEBULIN 7.
FT REPEAT 324 354 NEBULIN 8.
FT REPEAT 363 393 NEBULIN 9.
FT REPEAT 398 428 NEBULIN 10.
FT REPEAT 434 464 NEBULIN 11.
FT REPEAT 502 542 NEBULIN 12.
FT REPEAT 537 567 NEBULIN 13.
FT REPEAT 603 641 NEBULIN 14.
FT REPEAT 641 711 NEBULIN 15.
FT REPEAT 711 779 NEBULIN 16.
FT REPEAT 749 779 NEBULIN 17.
FT REPEAT 814 814 NEBULIN 18.
FT REPEAT 820 850 NEBULIN 19.
FT REPEAT 888 NEBULIN 20.
FT REPEAT 893 923 NEBULIN 21.
FT REPEAT 924 954 NEBULIN 22.
FT REPEAT 959 NEBULIN 23.
FT REPEAT 993 1023 NEBULIN 24.
FT REPEAT 1058 NEBULIN 25.
FT REPEAT 1064 1094 NEBULIN 26.
FT REPEAT 1102 1132 NEBULIN 27.
FT REPEAT 1137 1167 NEBULIN 28.
FT REPEAT 1168 1198 NEBULIN 29.
FT REPEAT 1204 1234 NEBULIN 30.
FT REPEAT 1237 1267 NEBULIN 31.
FT REPEAT 1272 1302 NEBULIN 32.
FT REPEAT 1308 1338 NEBULIN 33.
FT REPEAT 1346 1376 NEBULIN 34.
FT REPEAT 1381 1411 NEBULIN 35.
FT REPEAT 1412 1442 NEBULIN 36.
FT REPEAT 1448 1478 NEBULIN 37.
FT REPEAT 1481 1511 NEBULIN 38.
FT REPEAT 1516 1546 NEBULIN 39.
FT REPEAT 1552 1582 NEBULIN 40.
FT REPEAT 1590 1620 NEBULIN 41.
FT REPEAT 1625 1655 NEBULIN 42.
FT REPEAT 1656 1686 NEBULIN 43.
FT REPEAT 1692 1722 NEBULIN 44.
FT REPEAT 1725 1755 NEBULIN 45.
FT REPEAT 1760 1790 NEBULIN 46.
FT REPEAT 1796 1826 NEBULIN 47.
FT REPEAT 1834 1864 NEBULIN 48.
FT REPEAT 1869 1899 NEBULIN 49.
FT REPEAT 1900 1930 NEBULIN 50.
FT REPEAT 1936 1966 NEBULIN 51.
FT REPEAT 1969 1999 NEBULIN 52.
FT REPEAT 2004 2034 NEBULIN 53.
FT REPEAT 2040 2070 NEBULIN 54.
FT REPEAT 2078 2108 NEBULIN 55.
FT REPEAT 2113 2143 NEBULIN 56.
FT REPEAT 2144 2174 NEBULIN 57.
FT REPEAT 2180 2210 NEBULIN 58.
FT REPEAT 2213 2243 NEBULIN 59.
FT REPEAT 2248 2278 NEBULIN 60.
FT REPEAT 2284 2314 NEBULIN 61.
FT REPEAT 2322 2352 NEBULIN 62.
FT REPEAT 2357 2387 NEBULIN 63.
FT REPEAT 2488 2418 NEBULIN 64.
FT REPEAT 2423 2453 NEBULIN 65.
FT REPEAT 2456 2486 NEBULIN 66.
FT REPEAT 2491 2521 NEBULIN 67.
FT REPEAT 2527 2557 NEBULIN 68.
FT REPEAT 2565 2595 NEBULIN 69.
FT REPEAT 2600 2630 NEBULIN 70.
FT REPEAT 2631 2661 NEBULIN 71.
FT REPEAT 2666 2696 NEBULIN 72.
FT REPEAT 2699 2729 NEBULIN 73.
FT REPEAT 2734 2764 NEBULIN 74.
FT REPEAT 2770 2800 NEBULIN 75.
FT REPEAT 2808 2838 NEBULIN 76.
FT REPEAT 2843 2873 NEBULIN 77.
FT REPEAT 2874 2904 NEBULIN 78.
FT REPEAT 2909 2939 NEBULIN 79.
FT REPEAT 2942 2972 NEBULIN 80.
FT REPEAT 2977 3007 NEBULIN 81.
FT REPEAT 3013 3043 NEBULIN 82.
FT REPEAT 3051 3081 NEBULIN 83.
FT REPEAT 3086 3116 NEBULIN 84.
FT REPEAT 3117 3147 NEBULIN 85.
FT REPEAT 3152 3182 NEBULIN 86.

KW Transferase; DNA-directed DNA polymerase; DNA replication;
 F1 DNA-binding; DNA repair; Nuclear protein; Zinc finger; Polymorphism.
 F1 ZN_FING 3042 3057 C4 TYPE (POTENTIAL);
 F1 ZN_FING 3086 3104 C4 TYPE (POTENTIAL);
 F1 VARIANT 231 241 Q -> H
 FT FTID-VAR-008516,
 FT VARIANT 389 389 S -> T,
 FT VARIANT 1540 1540 /FTID-VAR-008517,
 FT VARIANT 2607 2607 /FTID-VAR-008518,
 FT VARIANT 2607 2607 /FTID-VAR-008519,
 FT CONFLICT 247 247 E -> Q (IN REF. 4 AND 5);
 FT CONFLICT 1156 1156 Y -> C (IN REF. 4 AND 5);
 SQ SEQUENCE 3130 AA: 352782 MW: 100700900F10HR14 "R664;

Query Match 5,18; Score 82.5; DB 1; Length 4140;
 Best Local Similarity 20.08; Pred. No: 62;
 Matches 62; Conservative 36; Mismatches 40; Indels 107; Gaps 13;

QY 31 VVCAKAOIPREELSVK PMVNTNTHINIQPQKQAFYRIENQIQEPLMSGLRIS 89
 DB 1577 VTSRKPRTRSTKQKIPKLKVSINLQ-----NSSQDNSVSDESIFES 1625
 QY 90 DKYLIAMVLAYFKRAGLYTSEYTMFTFVALYLANDMEDEEELYKEI - 148
 DB 1626 DP-----GPFSCYSLEISLSPIHNYNFNINIQGTGQCSF 1660
 QY 139 -----FWVA-IAGDSWRELEFQELRLRDFWAKNVIYAVSRRCQDEVMSSKPT 185
 DB 1661 YNGSOFVADNQLPKELSDVADQLFQGLATEKNEFLSHDNQK -CDEKRIHTD 1713
 QY 186 HWMLR-----DRMHHSQAMCYLRNEDDFP-----KGRDILPASQILFKAGVCD 243
 DB 1714 SASWIRSGTSPTEKSTIDSNENRNQKNSFIPLTTRSNSIMDSFCV-QQMEDCL 1771
 QY 234 SGGVSHNNSSSPEQEIF - - - - -HYTNREWSQEL - - - - -LMLPELL 270
 DB 1772 S-EKSLRNRSVSREVFSLSPQPNNSWIOCHTKRMGOSLSANSFATLSSPPGCLV 1830
 QY 271 DFECHDLHI 280
 DB 1841 DVAG-EDLEL 1849

RESULT 15
 DIM_ARATH STANDARD: PRT: 561 AA.
 AC Q9085; Q38808;
 DI 15-JUL-1998 (Ref. 36, Created)
 DI 01-MAR-2002 (Ref. 41, Last sequence update)
 DI 01-MAR-2002 (Ref. 41, Last annotation update)
 DE cell elongation protein DIMINUTO (cell elongation protein dwarf11).
 GN DIM OR DMF1 OR ATG19820 OR MPN9.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RX MEDLINE=95129837; PubMed=7828854;
 RA Takahashi T., Gasch A., Nishizawa N., Chua N.-H.;
 RT "The DIMINUTO gene of Arabidopsis is involved in regulating cell
 RT elongation.";
 RL Genes Dev. 9:97-107(1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RX Dukes B.P., Wu Y., Schulz B., Carlson T., Anna W., Feldman F.A.;
 RA Submitted (JUL-1994) to the EMBL/Genbank/DBP databases.

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RX MEDLINE=29277480; PubMed=10819429;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 4. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty B1 and TAC
 RT clones.";
 RL DNA Res. 7:141-145(2000).
 CC 1. FUNCTION: PLAYS A CRITICAL ROLE IN THE GENERAL PROCESS OF PLANT
 CC CELL ELONGATION.
 CC 2. SIMILARITY: BELONGS TO THE DIMINUTO FAMILY.
 CC
 CC THIS SWISS-Prot entry is copyright . It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, altered, by and for commercial
 CC entities requires a license agreement (www.ebi.ac.uk/seqdb/doc/seqdb/)
 CC or send an email to license@isb.scrib).

EMBL: L38520; AAA67055.1;
 DB EMBL: 012400; AAA20244.1;
 DB EMBL: AR025641; BAB01296.1;
 DB InterPro: IPR001575; oxid_FAB bind.
 DB Pfam: PF01565; FAD_binding_4; 1.
 FT CONFLICT 450 450 E -> G (IN REF. 1);
 FT CONFLICT 476 476 S -> L (IN REF. 1);
 FT CONFLICT 506 506 F -> L (IN REF. 1);
 FT CONFLICT 556 556 AYA -> PYP (IN REF. 1);
 SQ SEQUENCE 561 AA: 65094 MW: 70259769AM44299 (R664);

Query Match 5,18; Score 81.5; DB 1; Length 561;
 Best Local Similarity 28.88; Pred. No: 9.87;
 Matches 23; Conservative 14; Mismatches 18; Indels 27; Gaps 4;
 QY 245 PEQEIFH - - - - -YTNREWSQEL MLPELL DEFCT HDH 279
 DB 341 PIRVYHHRTRCLYWECKLLPFDQFWFYLLGLMLPPKVSLLKATGDAIKNYHMH 590
 QY 280 ILQEPVLEFDGTALEWHH 299
 DB 391 VIQMLVFLYKVGDALEWH 410

Search completed: July 31, 2002, 17:11:42
 Job time: 254 sec


```

QY 121 LYLANMEDEDEEYKYEIFFWALGDSWRELFPQFLRLRIGDFWAKMNYRAVVSRRKQCEVM 180
DB 121 LYLANMEDEDEEYKYEIFFWALGDSWRELFPQFLRLRIGDFWAKMNYRAVVSRRKQCEVM 180
QY 181 SKDPTHWAWLRDRPMHHSKAMRGYLRNEDOFFPPGPGLTPEASTLTHKACVDSGVSIN 240
DB 181 SKDPTHWAWLRDRPMHHSKAMRGYLRNEDOFFPPGPGLTPEASTLTHKACVDSGVSIN 240
QY 241 NSSSQEIEFYHYNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
DB 241 NSSSQEIEFYHYNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
RESULT 2
QYFWR2
ID QYFWR2 PRELIMINARY: PRT: 299 AA.
AC QYFWR2
DT 01-MAY-2000 (TREMBLERel. 14, Created)
DT 01-MAY-2000 (TREMBLERel. 14, Last sequence update)
DI 01-JUN-2001 (TREMBLERel. 17, Last annotation update)
DE P33 RINGO.
GN LS26/CLONE 5.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99396721; PubMed-10465793;
WA Forby L., Blazquez M., Palmer A., Britja R., Nebreda A.R.:
KT "A novel p34cdc2 binding and activating protein that is necessary and
RL sufficient to trigger G2/M progression in Xenopus oocytes."
RL Genes Dev. 13:2177-2189(1999).
DR EMBL: AJ249978; CAB58366.1;
DR INTERPRO: IPR000345; CYC home bind.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
SQ SEQUENCE: 299 AA; 34660 MW; C4701F109KCAZ0E1 CRG64;

```

```

Query Match: 97.7%; Score 1594.5; DB 13; Length 299;
Best local similarity: 98.3%; Pred. No. 9, 3e-156;
Matches 295; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 1 MRHMVSSTRASSIGSGVKVIGKGIPIHARVAVVAPKAAQIPEREELSVKPKMVENHNLQ 60
DB 1 MRHMVSSTRASSIGSGVKVIGKGIPIHARVAVVAPKAAQIPEREELSVKPKMVENHNLQ 60
QY 61 PQRQAAYRLEENFQIQTFFSMISCTKISIKYLIAMVLAYFKAAIYISETYIMNFFVA 120
DB 61 PQRQAAYRLEENFQIQTFFSMISCTKISIKYLIAMVLAYFKAAIYISETYIMNFFVA 119
QY 121 LYLANMEDEDEEYKYEIFFWALGDSWRELFPQFLRLRIGDFWAKMNYRAVVSRRKQCEVM 180
DB 121 LYLANMEDEDEEYKYEIFFWALGDSWRELFPQFLRLRIGDFWAKMNYRAVVSRRKQCEVM 179
QY 181 SKDPTHWAWLRDRPMHHSKAMRGYLRNEDOFFPPGPGLTPEASTLTHKACVDSGVSIN 240
DB 181 SKDPTHWAWLRDRPMHHSKAMRGYLRNEDOFFPPGPGLTPEASTLTHKACVDSGVSIN 239
QY 241 NSSSQEIEFYHYNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
DB 241 NSSSQEIEFYHYNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
RESULT 3
QYFWR2
ID QYFWR2 PRELIMINARY: PRT: 298 AA.
AC QYFWR2
DT 01-MAY-2000 (TREMBLERel. 14, Created)
DT 01-MAY-2000 (TREMBLERel. 14, Last sequence update)
DI 01-DEC-2001 (TREMBLERel. 19, Last annotation update)
DE P33 RINGO.

```

```

GN LS27.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99396721; PubMed-10465793;
WA Forby L., Blazquez M., Palmer A., Britja R., Nebreda A.R.:
KT "A novel p34cdc2 binding and activating protein that is necessary and
RL sufficient to trigger G2/M progression in Xenopus oocytes."
RL Genes Dev. 13:2177-2189(1999).
DR EMBL: AJ143500; CAB44296.1;
SQ SEQUENCE: 298 AA; 34573 MW; 88C752B64267BE76 CRG64;

```

```

Query Match: 89.5%; Score 1460; DB 13; Length 298;
Best local similarity: 89.0%; Pred. No. 2e-136;
Matches 267; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

```

```

QY 1 MRHMVSSTRASSIGSGVKVIGKGIPIHARVAVVAPKAAQIPEREELSVKPKMVENHNLQ 60
DB 1 MRHMVSSTRASSIGSGVKVIGKGIPIHARVAVVAPKAAQIPEREELSVKPKMVENHNLQ 60
QY 61 PQRQAAYRLEENFQIQTFFSMISCTKISIKYLIAMVLAYFKAAIYISETYIMNFFVA 120
DB 61 PQRQAAYRLEENFQIQTFFSMISCTKISIKYLIAMVLAYFKAAIYISETYIMNFFVA 119
QY 121 LYLANMEDEDEEYKYEIFFWALGDSWRELFPQFLRLRIGDFWAKMNYRAVVSRRKQCEVM 180
DB 121 LYLANMEDEDEEYKYEIFFWALGDSWRELFPQFLRLRIGDFWAKMNYRAVVSRRKQCEVM 179
QY 181 SKDPTHWAWLRDRPMHHSKAMRGYLRNEDOFFPPGPGLTPEASTLTHKACVDSGVSIN 240
DB 181 SKDPTHWAWLRDRPMHHSKAMRGYLRNEDOFFPPGPGLTPEASTLTHKACVDSGVSIN 239
QY 241 NSSSQEIEFYHYNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
DB 241 NSSSQEIEFYHYNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 298

```

```

RESULT 4
QYGLI
ID QYGLI PRELIMINARY: PRT: 298 AA.
AC QYGLI
DT 01-MAY-1999 (TREMBLERel. 10, Created)
DT 01-MAY-1999 (TREMBLERel. 10, Last sequence update)
DI 01-DEC-2001 (TREMBLERel. 19, Last annotation update)
DE SPEEDY PROTEIN.
GN SPY1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99219805; PubMed-10202150;
WA Leonard J.L., Bellinger R.W., Kudson R.L., Subramani S.:
KT "Speedy, a novel cyclin, is a regulator of the G2/M transition."
RL EMBO J. 18:1869-1877(1999).
DR EMBL: AJ143117; CAB394117.1;
DR INTERPRO: IPR000345; CYC home bind.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
SQ SEQUENCE: 298 AA; 34554 MW; 8870311995F41A7 CRG64;

```

```

Query Match: 88.7%; Score 1448; DB 13; Length 298;
Best local similarity: 88.0%; Pred. No. 6e-136;
Matches 264; Conservative 17; Mismatches 17; Indels 2; Gaps 2;

```


RX MEDLINE-21085660; PubMed-11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gajohori T., Bono H., Kasukawa T., Saito K.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Korchia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wanner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustineich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., King B., Ringwald M., Rodriguez L., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya I., Shibata Y., Storch K.,
 RA Watanabe-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuke S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK015441; BAB29849.1;
 DR MGD: MGI:1921923; 493045A17R1K.
 SQ SEQUENCE 268 AA; 30977 MW; 47ED87EFCABA2A4 C16*64;

Query Match 18.9%; Score 309; DB 11; Length 268;
 Best Local Similarity 41.5%; Pred. No. 1.9e-22;
 Matches 61; Conservative 46; Mismatches 48; Indels 2; Gaps 2;
 QY 47 VKPKWVNTLNLQPOEQAFYRLLENEQIQEFLSMDSLRISDKYLIAMVLAAYKRAAG 106
 DB 126 LEMKLRKRRVSTVRPFHVKVTKTETTFQVVFETLWIKMLKRSVLSKMLVAYTSK AG 178
 QY 167 LVSSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 166
 DB 179 LFSWQYPPHIEFLAYLIANDMEFNFNAPKQPIYFVLYKSYA,V PMHLLLEQVIESM3 247
 QY 167 YRAVVSRRCTDQVMSKQPTWAMLRDR 193
 DB 248 WKIWSRRECHETQAYNDLWVWARDR 264
 RESULT 8
 Q9D5G0
 ID Q9D5G0 PRELIMINARY: PRT: 173 AA.
 AC Q9D5G0;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DE 4930445A17R1K PROTEIN.
 GN 4930445A17R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;

RX MEDLINE-21085660; PubMed-11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gajohori T., Bono H., Kasukawa T., Saito K.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Korchia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wanner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustineich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., King B., Ringwald M., Rodriguez L., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya I., Shibata Y., Storch K.,
 RA Watanabe-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuke S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK015441; BAB29849.1;
 DR MGD: MGI:1921923; 4930445A17R1K.
 SQ SEQUENCE 173 AA; 19486 MW; FF41397E010B4411 C16*64;
 Query Match 6.2%; Score 102; DB 11; Length 173;
 Best Local Similarity 43.9%; Pred. No. 0.039;
 Matches 19; Conservative 18; Mismatches 19; Indels 0; Gaps 0;
 QY 47 VKPKWVNTLNLQPOEQAFYRLLENEQIQEFLSMDSLRISDKYLIAMVLAAYKRAAG 102
 DB 83 LEMKLRKRRVSTVRPFHVKVTKTETTFQVVFETLWIKMLKRSVLSKMLVAYTSK 106
 RESULT 9
 Q9DYS6
 ID Q9DYS6 PRELIMINARY: PRT: 1808 AA.
 AC Q9DYS6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOHETICAL 209.7 KDA PROTEIN.
 GN F17J16_140.
 OS Arabidopsis thaliana (Mouse ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons;
 OC eurosids 1; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=4702;
 RN 11
 RP SEQUENCE FROM N.A.
 RC D'Angelo M., Vozzi A., Modesto D., Piazzoli M., Valle G., Mowen H.W.,
 RA Radd S., Lemcke K., Mayer R.F.X., Goulet F., Salamant M.,
 RL Submitted (APR 2000) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC EU Arabidopsis sequencing project;
 RL Submitted (APR 2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A116327; CAB86946.1;
 DR InterPro: IPR003440; Glucan synthase.
 DR Pfam: PF02364; Glucan synthase; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1808 AA; 209729 MW; 0194701c705f7af C16*64;

Query Match 5.9%; Score 96; DB 10; Length 1808;
 Best Local Similarity 22.0%; Pred. No. 4;
 Matches 54; Conservative 26; Mismatches 88; Indels 74; Gaps 11;
 QY 45 LSVKPKWVNTLNLQPOEQAFYRLLENEQIQEFLSMDSLRISDKYLIAMVLAAYKRAAG 104
 DB 930 LSVKPKWVNTLNLQPOEQAFYRLLENEQIQEFLSMDSLRISDKYLIAMVLAAYKRAAG 982
 QY 105 AGLYTSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 147
 DB 983 LVSSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 1024
 QY 148 LVSSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 150
 DB 1024 LVSSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 1077
 QY 197 LVSSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 253
 DB 1078 LVSSEYTMNPFVALYLIANDMEDEDEYKYETTFEFLWALGDSWRELEFQELKRLNDEWAKWN 1129
 QY 254 N 254

Matches 68: Conservative 46; Mismatches 104; Indels 88; Gaps 16;

QY 7 VTRASSIGSGYKQVIGKHGHARVVGAR-KAGT-PEKEEL SVKPKMVNTHINLQPO 62
 DB 26 VAKRSKPPNCAQOQVRKRRNSPELPYRCKIOLPPKRELLRSVVIKELIAN HVPKDAI 84
 QY 63 ERQAFTRELLNE-----QIQEFLSDMSCLRSIDKYLIA--MVLAYFKKAGLYTSEYITM 115
 DB 85 DUEKXTRTFVGDRVIGWQFQAPLKWDKVLIQSLLHIVAGICLLIYDLRELNPYT---TIV 141
 QY 116 NFEVALYIANDMEDEDEEDYKYEFIPWALGDSWRELEFPQFLRI RDPFWAKMNYRAVVSRR- 174
 DB 142 STEVGVAG-----FVSTAQ-----AHPFWTHESYKANTVLSR 174
 QY 175 -CGDEVMSKDPITWAMLRDRPMHHSAMCYLRNDDFFPGKSLIPASCTILCHKAGV 241
 DB 175 ILMVGYVAGON-ILYDWMVDRHVRHK-----YSETDADE----- 208
 QY 232 CDSGGVSHNNSSSPQELFIHYTNREWSQELLMLP-PELLLDPECTHLLHILQELPLV 286
 DB 209 -----HNANKG-----FFSHVGW---LMLLKHPEVLRGRQTLMSDILLADIVV 249

RESULT 12

Q91AW0
 ID Q91AW0 PRELIMINARY: PRT: 1080 AA.
 AC Q91AW0:
 DT 01-DEC-2001 (Tremblrel, 19, Created)
 DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
 DE PUTATIVE FUSION PROTEIN P1 P2.
 OS beel chlorosis virus.
 OC Viruses; ssRNA positive strand viruses, no DNA stage; Lutovoviridae;
 CC Poliovirus.
 CC NCHI_TaxID=131082;
 RN [1]
 RC STRAIN: BGV-2A.
 RA Lemaitre G., Hauser S., Stevens M., Heve M.;
 RT "Biological properties and molecular characterization of beel
 chlorosis virus (BCV).";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF452024; AAK49956.1;
 SQ SEQUENCE 1080 AA; 121612 MW; CEF8B93D88F91E.2 CR654;

Query Match

Best Local Similarity 22.7%; Score 92; DB 12; Length 1080;
 Matches 72; Conservative 41; Mismatches 138; Indels 66; Gaps 16;

QY 11 SSICG -----SGVKQVIGKCHPHA-----RVVGAKKA Q:PEKEELSVKPKMVNTH 56
 DB 341 ASISCAVKKKASVSTDDGHSAPYISCKNVGHSQGVYRVNVSTLPIKIVNLT 400
 QY 57 LNLQPKQAFYRLLENEQIQEFLSDMSCLRSIDKYLIAWVLAYFKKAGLYTSEYITMN 116
 DB 401 POLVFETTAPOGRLEKEMDILLIYD---WSCQARALVKEPEK-GLVEGKGVK--- 452
 QY 117 FEVALYIANDMEDEDEEDYKYEFIPWALGDSW--RELFPQFLRI RDPFWAKM--YRAVV 171
 DB 453 -----SYDETSDRVSVPIVPPHLPKWKTERPPPPNPHI PFI -PESNARLWPELK 503
 QY 172 SRGCTGVMSKDPITWAMLRDRPMHHSAMCYLRNDDFFPGKSLIPASCTILCHKAGV 225
 DB 504 ARSCUDS-AKPFTHNASW-----NDDTDPPTDPTDTSHTCTFGS 544
 QY 226 CHKAGVDSGGVSHNNSSSPQELFIHYTNREWSQELLMLP-PELLLDPECTHLLHILQ 282
 DB 545 KHESCI-DCSIPFNKSRREYRE SCWTDSEKKRAIKTQAKLEKKEFGGVKWCNCIT 602
 QY 283 EPLVGLDFDCTALEMHH 299
 DB 603 QEVGTEEGSLPKFYH 619

RESULT 14

Q9B530
 ID Q9B530 PRELIMINARY: PRT: 1624 AA.
 AC Q9B530:
 DT 01-AUG-1998 (Tremblrel, 07, Created)
 DT 01-AUG-1998 (Tremblrel, 07, Last sequence update)
 DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
 DE 1624AA LONG HYPOTHETICAL REVERSE GYRASE.
 GN PH0860.
 OS Pyrococcus horikoshii.
 CC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 CC NCHI_TaxID=54954;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN 014;
 RX MF01NF-08444137; PubMed 9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Baikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S., Kusunagi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohkubo Y.,
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Koshida N., Imachi A.,
 RA Aoki K., Yoshizawa T., Nakamura Y., Robb F.J., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii of 4.°C."
 RL DNA RES. 5:55-76(1998).
 DR EMBL: AP000004; XAA29893.1;
 DR HSSP: 006612; 1671.
 DR InterPro: IPR001410; 16AD.
 DR InterPro: IPR004601; DNATop1 ATP bind.
 DR InterPro: IPR004602; DNATop1 DNA bind.
 DR InterPro: IPR004586; HININ.
 DR InterPro: IPR004587; HININ.
 DR InterPro: IPR000688; HYP.
 DR InterPro: IPR002204; Intein.
 DR InterPro: IPR004042; Intein endonuc.
 DR InterPro: IPR000480; Pro topoisomerase.
 DR InterPro: IPR002936; Toprim.
 DR Pfam: PF01751; Toprim; 4.
 DR PRINTS: PR00417; PRF1PSMEASE1.
 DR SMART: SM00487; DEXDc 1.
 DR SMART: SM00405; HININ; 1.
 DR SMART: SM00406; HININ; 1.
 DR SMART: SM00437; Top1Ac; 1.
 DR SMART: SM00436; Top1Bc; 1.
 DR SMART: SM00493; TOPRIM; 1.
 DR PROSITE: PS00041; PROTEIN SPLITTING; DNKN.WN 1.
 KW Complete proteins.
 SQ SEQUENCE 1624 AA; 187068 MW; 1B97FEF42F10EFF5 CR654;

Query Match

Best Local Similarity 19.6%; Score 91; DB 17; Length 1624;
 Matches 46; Conservative 44; Mismatches 75; Indels 70; Gaps 12;

QY 88 ISUKYLLIAMVLAYFKKAGLYTSEYITMNFEVALYI ANIMDEDEEDYKY 146
 DB 1177 LREYINAMIGYED-ASGTEFLRAVLTSKRLDLMLSVLYQLGVNLLDPEHIAWV 1240
 QY 137 EITFWALGDSWRELEFPQFLRI RDPFWAKMNYRAVVSRR- 176
 DB 1236 ELLELDL EFEREKIYVIGKAKKASQDKVYSKKNAGLEFVASTERKLEKKEFGKKNK 1294
 QY 177 DFVMSKDPITWAMLRDRPMHHSAMCYLRNDDFFPGKSLIPASCTILCHKAGV 225
 DB 1295 DEELPREVAKVLEYAHSPEKFLNSLVEAKVWVRKEITERYIKLYITITITIN 1354
 QY 212 EPRGPGITFACCTIICHKAGVDSGGVSHNNSSSPQELFIHYTNREWSQELLMLP 286
 DB 1455 ISN-GLVSHINITY HKT 165THVSNGLGVAKVYITLQELGEEYEDPKWCEE 1400

```

X NBL_TaxID=7227.
X [-]
X
X SEQUENCE FROM N.A.
X STRAIN=BERKELEY.
X MEDLINE=20196006; PubMed=10731132;
X
X Adams M.D., Geluker S.E., Holt R.A., Evans C.A., Gorayeb J.D.,
X Ananidis P.G., Scherer S.E., Li P.W., Hoskins P.A., Valle R.F.,
X George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
X Stanton G.C., Wortman J.K., Tanciel M.D., Zhang Q., Chen L.X.,
X Hudson R.G., Rogers J.H., Blazek R.G., Champ C., Pfeiffer B.D.,
X Wan K.H., Doyle C., Baxter F.G., Holt G., Nelson C.W., McKus G.L.G.,
X Arif J.P., Aghayani A., An H.-J., Andrews-PlankKoch C., Baldwin D.,
X Bailey P.M., Basu A., Baxendale T., Bayraktaroglu I., Gearley E.M.,
X Beeson K.Y., Benos P.V., Berman P.P., Bhattacharya P., Bolshakoy S.,
X Borkova D., Botchan M.K., Bouck J., Brokstein P., Brothier P.,
X Burris K.C., Husam D.A., Butler H., Cadieu E., Center A., Chandra I.,
X Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
X de Pablo B., Delcher A., Deng Z., Mays A.L., Dew I., Dietz S.M.,
X Davidson K., Dong L.E., Dewes M., Duan Esch S., Durkin B.C., Dunn P.,
X Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
X Fouts C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
X Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
X Harris N.L., Harvey D., Heiman T.J., Hernandez J.K., Houck J.,
X Hostin D., Houston K.A., Howland T.J., Wei M.-H., Hsueh C.,
X Jalili M., Kalish P., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
X Kimmel B.E., Kodira C.D., Kraft C., Krahl S., Kulp D., Lai Z.,
X Laske P., Lei Y., Levitsky A.A., Li G., Li G., Liang Y., Lin X.,
X Liu X., Mattel B., McIntosh T.C., McLeod M.P., Moperson D.,
X Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostrofi A.,
X Nelson S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
X Nelson D.F., Nelson K.A., Nizkor E., Nusslein E.F., Padgett J.M.,
X Palazzolo M., Plattman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
X Reiner K., Remington K., Saunders R.B.C., Scheeler F., Shen H.,
X Shue R.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith I.,
X Spier E., Spradling A.C., Stapleton M., Strong K., Sun E.,
X Stairskas P., Teeter C., Turner P., Ventler E., Wang A.H., Wang X.,
X Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
X Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
X Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao G., Zheng L.,
X Zhang X.H., Zhang F.N., Zheng W., Zhou X., Zhu S., Zhu X., Smith H.O.,
X Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;
X "the genome sequence of Drosophila melanogaster.";
X Science 287:2185-2195(2000).
X
X EMBL: AE003708; AF595149.1;
X FlyBase: FBgn0048308; CG5635.
X InterPro: IPR00152; FATC.
X InterPro: IPR000403; P13_P14_kinase.
X Pfam: PF02260; FATC; 1.
X Pfam: PF00454; P13_P14_kinase; 1.
X SMART: SM00146; PI3K; 1.
X PROSITE: PS00915; P13_4_KINASE_1; 1.
X PROSITE: PS00916; P13_4_KINASE_2; 1.
X PROSITE: PS0290; P13_4_KINASE_3; 1.
X SEQUENCE 4429 AA; 27866 MW; AE0E440544FAD0DB CRC64;
X
X Query Match 5.48; Score 88.5; OH 5; Length 2429;
X Best local Similarity 20.48; Pred. Num. 24;
X Matches 66; Conservative 47; Mismatches 118; Indels 93; Gaps
X
X 51 MVRNTHNLQPGQGAIVPLENQIGF-----LSMD--SCLYISIKYLLAMV 97
X | | | | | | | | | | | | | | | | | | | | | | | | | | |
X 1227 MEKYNKLIFAKASIASIFIT-SNKLIPPIHW-SILPAI-SLSYEPWM-AVY-VFVMECW 1286
X | | | | | | | | | | | | | | | | | | | | | | | | | | |
X
X 98 LAYPKRAAGLYTSYNTMNFVALIANDMEHQEF-----DYKYEIPWALGJSWRELP 152
X | | | | | | | | | | | | | | | | | | | | | | | | | | |
X
X 1287 LGFDKLAATSPAFAKTLLPPIKLLNSLENVESLSMLDYTFEGETSSIAFNSQEIER 1346
X | | | | | | | | | | | | | | | | | | | | | | | | | | |
X
X 153 -----OFLRLRDDPFAKWNKAVV--SKRC-----CQVNSKDP 184
X | | | | | | | | | | | | | | | | | | | | | | | | | | |
X
X 1347 NKRAIKKELHLCYETVIRFNNNIPINISNVVMASNHCCALYLSIMYLEDWAL-----SSEP 1402
X | | | | | | | | | | | | | | | | | | | | | | | | | | |
X
X 195 THWAWLEPPPHHNSKAPGY-----LSNECEETPPGPEL-----SSEP 230
X | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
Db 1403 KSKADFLONECFQDGAKKAYESIGCLDAIPGVNPMKSRLLDTLGGNSNLSTHLLSDHLD 1462
QY 221 -ASSTIC--- HKACVDSGCVSHNNSSDPQEHFYINKEWSQEHMPPHLLDDPECT 275
Db 1463 KASGQCIDIMKGNGLNSFAKIQHQHLEPDYEIF-WRLQW-- -USLIDPKHQ 1512
QY 276 HLLHLLQEPLVGLPEPDGTALEMHH 299
Db 1513 ONOTVVRTSU-DLEQE- -FKRRH 1532
```

Search completed: July 31, 2002, 17:10:51
Job time: 268 sec

GenBank version 1.5
Copyright (c) 1994-2000 Empagen, Ltd.

OM protein protein search, using sw model

Run On: July 31 2002 17:11:55 Search time 47.09 seconds
(with all alignments)
612.294 Million cell updates/sec

Title: US-09-889-592-2
Perfect score: 1632
Sequence: 1 MPMQVTPASSHQQWQ
Scoring table: BLOSUM62
Gapop 10.0, Gapext 0

Searched: 28338 seqs, 3608634 residues
Total number of hits satisfying cluster parameters: 28338

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR71:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	PIR	IF	Description
1	309.5	19.0	308	2	T46338	hypothetical prote
2	95	5.9	1808	2	T4792	hypothetical prote
3	94.5	5.7	563	2	F8788	probable D-lactate
4	91	5.6	1497	2	S7250	sex-determining tr
5	91	5.6	1624	2	G7129	probable reverse a
6	88.5	5.4	532	2	H7389	NADH dehydrogenase
7	88	5.4	473	2	T4704	probable oxoglutar
8	87.5	5.4	320	2	A7749	abortive phage res
9	87.5	5.4	564	2	D8134	D-lactate dehydrog
10	87	5.3	346	2	T2047	hypothetical prote
11	87	5.3	1027	2	T6144	oxoglutarate dehyd
12	86	5.3	416	2	A7370	hydrogenase large
13	86	5.3	1480	2	D8074	cystic fibrosis tr
14	85.5	5.2	1783	2	A7201	serine/threonine k
15	84.5	5.2	780	2	G7174	probable penicilli
16	84	5.1	586	2	G7428	hypothetical prote
17	84	5.1	1009	2	T4783	probable oxoglutar
18	84	5.1	1468	1	S4818	hypothetical prote
19	83.5	5.1	265	2	J07515	beta 2 toxin - Clo
20	83.5	5.1	507	2	H2580	glycerol-3-phospha
21	83	5.1	268	2	A7745	probable NAM (no a
22	83	5.1	1208	2	T4773	phosphatidic acid
23	83	5.1	6669	2	S9124	nebulin, skeletal
24	82	5.0	2526	2	T21399	myosin RhoGAP prot
25	81.5	5.0	344	2	A7340	hypothetical prote
26	81.5	5.0	561	2	G7138	hwarf protein A
27	81.5	5.0	635	2	S9125	hypothetical prote
28	81.5	5.0	1291	2	T07692	hypothetical prote
29	81	5.0	225	1	G738	C-reactive protein

ALIGNMENTS

RESULT 1

T46338
hypothetical protein DKFZp434A1014.1 - human (fragment)

Species: Homo sapiens (man)
Date: 04-Feb-2000 #sequence_revision: 04-Feb-2000 #text_change: 04-Feb-2000
Accession: T46338
Submitted to the Protein Sequence Database, January 2000
Submitted by: K. Royer, A. Mewes, H. W. Gassenhuber, J. Wiemann, S.
Reference number: 223037
Accession: T46338
Status: preliminary
Molecule type: mRNA
Residues: 1-308 <AAA>
Cross-references: EMBL:AL137266
Experimental source: adult testis; clone DKFZp434A1014
Notes: DKFZp434A1014.1

Query Match 19.0% Score 309.5; ID 2; Length 308;
Host Local Similarity 34.0% Pred No 3.0e 20;
Matches 75; Conservative 29; Mismatches 54; Indels 57; Gaps 4;

```

37 35 PKAQTPRRRL-----SVKPKMVRNTHINLOPHQAPYRIENPQIOPFLSMDSCLR 87
   || || || | | | | | | | | | | | | | | | | | | | | | | | |
61 61 KAVLAPFPEELVWAEMLGEGIKMKIKKPPFVSLVLEPHEAFNKLIEPVIRFLAWKULP 150
   || || || || || || || || || || || || || || || || || || ||
68 88 ISDKYLIAVLAVFKPAAGLYTSEYTTMNFVVALYLIANTMEEDFEDYKVIFFWALGDS- 146
   || || || || || || || || || || || || || || || || || || ||
151 151 VSURYLIAVLAVFSK-AGFISWQYQKLFALYLIANTMEEDFEDYKVIFFWALGDS 209
   || || || || || || || || || || || || || || || || || || ||
17 147 -----WRLFPQ-----FRLR 158
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
210 210 SPFLLEKRPFGVFSNMFAKPNFSLPIVVPFPEFLEKPMNFAFFNDELVLVLEKDP 269
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
159 159 DDFWAKNNYKAVYSKRCDEVMKSDPTHWALDR 193
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
270 270 FHFPCSCRAWSPDELEETQAYDPRHWVWARD 304
   || | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 2

hypothetical protein F7J16.150 - Arabidopsis thaliana
Species: Arabidopsis thaliana (mouse-ear cress)

Date: 20-Apr-2000 #sequence_revision: 20-Apr-2000 #text_change: 20-Apr-2000
Accession: T47792
Submitted to the Protein Sequence Database, April 2000
Submitted by: M. Vezzi, A. Mewes, D. Eidezz, M. Valle, G. Mewes, H.W. Rudd, S.
Reference number: 224476
Accession: T47792
Status: preliminary
Molecule type: DNA

Search completed: July 3, 2002, 11:09:11
Job time: 4 hr 50c

```

? SEQ ID No 4
? LENGTH: 1474
? TYPE: PRT
? ORGANISM: Homo sapiens
US 09-174 077-4

Query Match: 5.3% Score 86; ID# 4; Length 1474;
Best Local Similarity 20.7% Prev. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 0

QY 5 KASSIGSSVQGV --EGHHHAPVVAPRAQIPIREI SVKIKVKNTHI
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 220 QASACGCGEITVIAGTQAGIGRMKMYRQRAKISEK LVISIMENIQNS
64 KATVKELENAQIGELISMISLRISEKYLIVMLAVIEKAAVLT
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 64 KATVKELENAQIGELISMISLRISEKYLIVMLAVIEKAAVLT
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 278 DEAMKMLENRQIE ---LKLTK AAVVYFNSSAFFEYDFEYVFLIN
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 110 ----- SEYTMNPFVALTIANMEDEHRYKDFPWALGDSNDFEYDFEYVFLIN
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 328 LKGIILKRIETTSFGLVLENAVTRQ - EFWAV QIWDNLGALNK
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 164 AKMYRAVVSRRCCDEVMSKDP 1R5
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 376 QKQYKILEYNTLIEVMENVI 494
111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 4
US 07-647-621 2
? Sequence 2, Alignment: 86/976,47621
? Patent No. 5407796
? GENERAL INFORMATION:
? APPLICANT: cutting, dairy

```

TITLE OF INVENTION: CYSTIC FIBROSIS MUTATION CLUSTER
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, McKie and Lockett
 STREET: 1001 G Street, N.W.
 CITY: Washington, D.C.
 COUNTRY: USA
 ZIP: 20001
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS DOS
 SOFTWARE: Patent In Release #1-0, Version #1-25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/64,621
 FILING DATE: 19910104
 CLASSIFICATION: 4-5
 ATTORNEY/AGENT INFORMATION:
 NAME: kadan, sarah d
 REGISTRATION NUMBER: 42,141
 REFERENCE/BOOKET NUMBER: 1107,040010
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508 9100
 TELEFAX: 202-508-9100
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1480 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: HOMO SAPIENS
 DIS-07-637-621 2

Query Match	5.3%	Score 86;	DB 1;	Length: 1480;
Best Local Similarity	20.7%	Prod. No. 2;		

Matches 42: Conservative 3 M. matches 74: Indels 50: Gaps 8:

QY 9 KASSICSGVKQV-----GKCHAKAV AKAL DEFEFISVKMMKVPNHNILQPOF 63
 DB 220 QASAPCGHGLIVAI PQAGHMMKYL QAGKSER-IVTSEMENISVKAYCW 277
 QY 64 ROAFYRLLENEQIFELSMDSCTIRISKYLIAMVLAIFKRAAGLYT-----109
 DB 278 EEMAKMIENLRQTE-----LKLTRK AAYVYNSSAFESFEFVVLVLPVAL 327
 QY 110 -----SYITMNFVALYLANIMEHEDEYKYLFIHWAIGDSWRHLFQDFLRDDPW 162
 DB 428 IKGILKFIETTSFIVLPMVITQ-----EPWAV-QTWYDSGAINKIQ-DFL 375

QY 163 ARMYPAVVSPECDDEVMSKDEPT 185
 DB 376 QKQYKTIENLYNLTTEVVMNVI 398

RESULT 6
 US-08-135-809A-2
 Sequence 2, Application US/08135809A
 Patent No. 5686677
 GENERAL INFORMATION:
 APPLICANT: CHENG, SENG H.
 APPLICANT: DITULLIO, PAUL
 APPLICANT: EBERT, KARL M.
 APPLICANT: MEADE, HARRY M.
 APPLICANT: SMITH, ALAN E.
 TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS CONTAINING
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESS: GENZYME CORPORATION
 STREET: ONE MOUNTAIN ROAD
 CITY: FRAMINGHAM
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/135,809A
 FILING DATE: 13-OCT-1993
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: LASSEN, ELIZABETH
 REGISTRATION NUMBER: 31,845
 REFERENCE/DOCKET NUMBER: 104-9,12
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1480 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-135-809A-2

Query Match 5.3%, Score 86; DB 1, length 1480;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 42: Conservative 37 Mismatches 74: Indels 50: Gaps 8;

QY 9 KASSICSGVKQV-----GKCHAKAVVVGAKKALGPEELSVPKPMVRNTHNLQPOF 63
 DB 220 QASAPCGHGLIVAI PQAGHMMKYLQAGKSER-IVTSEMENISVKAYCW 277
 QY 64 ROAFYRLLENEQIFELSMDSCTIRISKYLIAMVLAIFKRAAGLYT-----109
 DB 278 EEMAKMIENLRQTE-----LKLTRK-----AAYVYNSSAFESFEFVVLVLPVAL 327
 QY 110 -----SYITMNFVALYLANIMEHEDEYKYLFIHWAIGDSWRHLFQDFLRDDPW 162
 DB 428 IKGILKFIETTSFIVLPMVITQ-----EPWAV-QTWYDSGAINKIQ-DFL 375

Matches 42: Conservative 3 M. matches 74: Indels 50: Gaps 8:

QY 9 KASSICSGVKQV-----GKCHAKAV AKAL DEFEFISVKMMKVPNHNILQPOF 63
 DB 220 QASAPCGHGLIVAI PQAGHMMKYL QAGKSER-IVTSEMENISVKAYCW 277
 QY 64 ROAFYRLLENEQIFELSMDSCTIRISKYLIAMVLAIFKRAAGLYT-----109
 DB 278 EEMAKMIENLRQTE-----LKLTRK AAYVYNSSAFESFEFVVLVLPVAL 327
 QY 110 -----SYITMNFVALYLANIMEHEDEYKYLFIHWAIGDSWRHLFQDFLRDDPW 162
 DB 428 IKGILKFIETTSFIVLPMVITQ-----EPWAV-QTWYDSGAINKIQ-DFL 375

QY 163 ARMYPAVVSPECDDEVMSKDEPT 185
 DB 376 QKQYKTIENLYNLTTEVVMNVI 398

RESULT 5
 US-08-146-742A-2
 Sequence 2, Application US/ 814,42A
 Patent No. 5670488
 GENERAL INFORMATION:
 APPLICANT: Gregory, R.J., Attorney, B.J. Couture, L.A., Smith,
 APPLICANT: A.E.
 TITLE OF INVENTION: GENETICALLY DERIVED CYCLIC PEPTIDES
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BRUMBAUGH, GRAVES, I WHITE & RAYMOND
 STREET: 30 ROCKEFELLER PLAZA
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US 07/145,413
 FILING DATE: 02-DEC-1992
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Selde, Rochelle K.
 REGISTRATION NUMBER: 23,300
 REFERENCE/DOCKET NUMBER: A 669 Contingent F.P. 124-9,113
 TELEPHONE: (212) 408-2500
 TELEFAX: (212) 765-2519
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1480 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-146-742A-2

Query Match 5.3%; Score 86; DB 1, length 1480;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 42: Conservative 37 Mismatches 74: Indels 50: Gaps 8;

QY 9 KASSICSGVKQV-----GKCHAKAVVVGAKKALGPEELSVPKPMVRNTHNLQPOF 63
 DB 220 QASAPCGHGLIVAI PQAGHMMKYLQAGKSER-IVTSEMENISVKAYCW 277
 QY 64 ROAFYRLLENEQIFELSMDSCTIRISKYLIAMVLAIFKRAAGLYT-----109

QY 163 AKMYKAVVSRCCDVMNSKPT 185
DB 376 QKQYKTYLNTTTEVVMNVT 398

RESULT 7
US-08-466-886-17
: Sequence 17, Application US/08466886
: Patent No. 5776677
: GENERAL INFORMATION:
: APPLICANT: Tsui, Lap Chee
: APPLICANT: Riordan, John R.
: APPLICANT: Rommens, Johanna M.
: APPLICANT: Kerem, Bal-Sheva
: APPLICANT: Collins, Francis S.
: APPLICANT: Iannuzzi, Michael C.
: APPLICANT: Drumm, Mitchell L.
: APPLICANT: Buckwald, Manuel
: TITLE OF INVENTION: Cystic Fibrosis Gene
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX
: STREET: 1100 New York Avenue, N.W.
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.40
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,886
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldstein, Jorge A.
: REGISTRATION NUMBER: 29,021
: REFERENCE/DOCKET NUMBER: 1329,0010006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1480 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08 466-886-17

Query Match 5.3% Score 86; DB 1; Length 1480;
Best Local Similarity 20.7%; Pred. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;
QY 9 RASSIGSGVKQVI-----GKGHHARVVGAKKAGIPEREELSVAKMYVNTNHLNLPQW 63
DB 220 GASAFGLGLVIALFVAGLGMMMKYRQRAKISR LVTSSEMTENTQSVKAYTW 277
QY 64 KAAYRIENFQIEFLMSDGLRISDKYLAMVLAYEKRAKLYT ----- 109
DB 278 KEAMKMEINLRQIE-----LKLTK-----AAVYKFNSSAFPSGPFVVLVSLVVAL 327
QY 110 -----SEYTMFFVVALYANIMDEEDRYAYEIPWALGDSWRELPQFLRLRDDFW 162
DB 328 IKGILRKIKFTTISFCVLRVAVTRQ-----FPWAV-QTWYDLSGLINKIQ-DLF 375
QY 163 AKMYKAVVSRCCDVMNSKPT 185
DB 376 QKQYKTYLNTTTEVVMNVT 398

RESULT 8
US-08-951-912-2
: Sequence 2, Application US/08951912
: Patent No. 5972995
: GENERAL INFORMATION:
: APPLICANT: Fischer, Horst
: APPLICANT: Illek, Beate
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYCLIC
: TITLE OF INVENTION: FIBROSIS THERAPY
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6400 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.40
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: us/08/951,912
: FILING DATE: 16-OCT-1997
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 41,492
: REFERENCE/DOCKET NUMBER: 200116,403
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622 4900
: TELEFAX: (206) 682 6041
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1480 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08 951-912-2

Query Match 5.3% Score 86; DB 2; Length 1480;
Best Local Similarity 20.7%; Pred. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;
QY 9 RASSIGSGVKQVI-----GKGHHARVVGAKKAGIPEREELSVAKMYVNTNHLNLPQW 63
DB 220 GASAFGLGLVIALFVAGLGMMMKYRQRAKISR LVTSSEMTENTQSVKAYTW 277
QY 64 KAAYRIENFQIEFLMSDGLRISDKYLAMVLAYEKRAKLYT ----- 109
DB 278 KEAMKMEINLRQIE-----LKLTK-----AAVYKFNSSAFPSGPFVVLVSLVVAL 327
QY 110 -----SEYTMFFVVALYANIMDEEDRYAYEIPWALGDSWRELPQFLRLRDDFW 162
DB 328 IKGILRKIKFTTISFCVLRVAVTRQ-----FPWAV-QTWYDLSGLINKIQ-DLF 375
QY 163 AKMYKAVVSRCCDVMNSKPT 185
DB 376 QKQYKTYLNTTTEVVMNVT 398

RESULT 9
US-08-951-912-6
: Sequence 6, Application US/08951912
: Patent No. 5972995
: GENERAL INFORMATION:
: APPLICANT: Fischer, Horst
: APPLICANT: Illek, Beate
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYCLIC
: TITLE OF INVENTION: FIBROSIS THERAPY
: NUMBER OF SEQUENCES: 6

1 CORRESPONDENCE ADDRESS:
 2 ADDRESS: SEED and BERRY P
 3 STREET: 6300 Columbia Center, 7th Fl in Ave. A
 4 CITY: Seattle
 5 STATE: Washington
 6 COUNTRY: USA
 7 ZIP: 98104
 8
 9 COMPUTER READABLE FORM:
 10 MEDIUM TYPE: floppy disk
 11 COMPUTER: IBM PC compatible
 12 OPERATING SYSTEM: PC DOS/MS DOS
 13 SOFTWARE: Patent In Release 4.0 Version #1.0
 14
 15 CURRENT APPLICATION DATA:
 16 FILING DATE: 16-CT-1997
 17 APPLICATION NUMBER: US/08-461-4
 18 CLASSIFICATION: 61A
 19 ATTORNEY/AGENT INFORMATION:
 20 NAME: Maki, David J.
 21 REGISTRATION NUMBER: 31,394
 22 REFERENCE/DOCKET NUMBER: 20016,1003
 23 TELECOMMUNICATION INFORMATION
 24 TELEPHONE: (206) 622-4400
 25 TELEFAX: (206) 692-6041
 26
 27 INFORMATION FOR SEQ ID NO: 6:
 28 SEQUENCE CHARACTERISTICS:
 29 LENGTH: 1480 amino acids
 30 TYPE: amino acid
 31 STRANDEDNESS:
 32 TOPOLOGY: linear
 33
 34 US-08-461-461-4

Query Match 5.3%, Score 86; DB 2; Length 1480;
 Best Local Similarity 20.7%, Fred. No. 2;
 Matches 42; Conservative 37; Mismatches 74; Indels 50; Gaps 8;

1 9 RASSICGSGVKQVI-----CKGHPIHARVVGAKKAOIPERELSVKPKMVRNTHINLIQPOE 63
 2 220 QASAPQRIQFLIVLALFQARIGPMKKYPTGPAKRISEF--LVITSEMIENT--SVKAYCW 277
 3 64 RQAFYRLLENQIQEFLSMDSCLRIISKYLIAMVLAYFKRAAGLYT----- 109
 4 278 RFAEMKMIENLPQRI-----LKLTRK- AAVRVYFNSSAFFPSGFFVFLSVLYAL 327
 5 110 -----SEYTTMNFVVALYLANDMEDEEDYKYEIFPWALGDSWRELFPPQLRDEFW 162
 6 328 IKGIIILPKIITTSFCIVLPMATPQ-----FPWAV-QTWYNSLGAINKIQ-DFL 375
 7 163 AKMYRAVVSRRCCDEVMKSDIPT 185
 8 376 QKQEKYLEYNLTTEVVMENVT 398

RESULT 10
 US-08-461-461-4
 Sequence 2, Application US/0846414
 Patent No. 5981178
 GENERAL INFORMATION:
 APPLICANT: Tsui, Lap-chee
 APPLICANT: Kerem, Bat-Sheva
 TITLE OF INVENTION: Introns and Exons of the Cystic Fibrosis Gene and
 FILE REFERENCE: 3477-61, 03477/13984
 CURRENT APPLICATION NUMBER: US/08-461-461B
 CURRENT FILING DATE: 1995-06-06
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 2
 LENGTH: 1480
 TYPE: PRI
 ORGANISM: Homo sapiens
 APPLICANT: and Rafter, Patrick W.
 US-08-461-461-4

Query Match 5.3%, Score 86; DB 2; Length 1480;
 Best Local Similarity 20.7%, Fred. No. 2;
 Matches 42; Conservative 37; Mismatches 74; Indels 50; Gaps 8;

1 9 RASSICGSGVKQVI-----CKGHPIHARVVGAKKAOIPERELSVKPKMVRNTHINLIQPOE 63
 2 220 QASAPQRIQFLIVLALFQARIGPMKKYPTGPAKRISEF--LVITSEMIENT--SVKAYCW 277
 3 64 RQAFYRLLENQIQEFLSMDSCLRIISKYLIAMVLAYFKRAAGLYT----- 109
 4 278 RFAEMKMIENLPQRI-----LKLTRK- AAVRVYFNSSAFFPSGFFVFLSVLYAL 327
 5 110 -----SEYTTMNFVVALYLANDMEDEEDYKYEIFPWALGDSWRELFPPQLRDEFW 162
 6 328 IKGIIILPKIITTSFCIVLPMATPQ-----FPWAV-QTWYNSLGAINKIQ-DFL 375
 7 163 AKMYRAVVSRRCCDEVMKSDIPT 185
 8 376 QKQEKYLEYNLTTEVVMENVT 398

Query Match 5.3%, Score 86; DB 2; Length 1480;
 Best Local Similarity 20.7%, Fred. No. 2;
 Matches 42; Conservative 37; Mismatches 74; Indels 50; Gaps 8;

1 9 RASSICGSGVKQVI-----CKGHPIHARVVGAKKAOIPERELSVKPKMVRNTHINLIQPOE 63
 2 220 QASAPQRIQFLIVLALFQARIGPMKKYPTGPAKRISEF--LVITSEMIENT--SVKAYCW 277
 3 64 RQAFYRLLENQIQEFLSMDSCLRIISKYLIAMVLAYFKRAAGLYT----- 109
 4 278 RFAEMKMIENLPQRI-----LKLTRK- AAVRVYFNSSAFFPSGFFVFLSVLYAL 327
 5 110 -----SEYTTMNFVVALYLANDMEDEEDYKYEIFPWALGDSWRELFPPQLRDEFW 162
 6 328 IKGIIILPKIITTSFCIVLPMATPQ-----FPWAV-QTWYNSLGAINKIQ-DFL 375
 7 163 AKMYRAVVSRRCCDEVMKSDIPT 185
 8 376 QKQEKYLEYNLTTEVVMENVT 398

RESULT 12
 US-08-461-605-2
 Sequence 2, Application US/084641605
 Patent No. 5981714
 GENERAL INFORMATION:
 APPLICANT: Cheng, Song H., Marshall, John, Gregory, Richard J.
 APPLICANT: and Rafter, Patrick W.

TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR CYSTIC FIBROSIS
 TITLE OF INVENTION: TRANSMEMBRANE CONDUCTANCE REGULATOR AND USES
 TITLE OF INVENTION: THEREFOR
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & CUCKFIELD
 STREET: 60 STATE STREET, SUITE 510
 CITY: BOSTON
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/691,605

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/114,950

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Hanley, Elizabeth A.

REGISTRATION NUMBER: 33,505

REFERENCE/DOCKET NUMBER: NZ1-029

TELEPHONE: (617) 227-7400

TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1480 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-691-605-2

Query Match 5.48; Score 86; DB 2; Length 1480;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICSGSVKQVI-----GKHPHARVVCARKAKQIPEPEELSVKPKVNRNTHLNPQE 63

DB 220 CASAFCCGFLVIALFQA:G:GMMMKYINQAGKISER LVITSEMENIQSVKAYCW 277

QY 64 RQAFYRLLENEQTEFLSMDSCLRSISKYLLIAMVLAYFKRAAGLYI----- 109

DB 278 EAMKMKIENLRQTE-----LKLTRK---AAYVYFNSSAFFSGFEVVLVLPYAL 327

QY 110 -----SEYITNNFVVALYLANIMEDEEDYKYIEFWALGDSWRFQDFPIRPPFW 162

DB 328 IKGILKRIPTTISFQVILRMATRO-----FWAVQTWYDSLGAIKNTQ-DFL 375

QY 163 AKMYRAVVSRRCCDVMSKDP 185

DB 376 QKQYKILEYNIETTVVMEVNI 398

RESULT 13

US-08 455-552A-14

Sequence 14, Application US/08455552A

Patent No. 5990279

GENERAL INFORMATION:

APPLICANT: Carter, Barrie J.

APPLICANT: Flotte, Terence

APPLICANT: Afione, Sandra

APPLICANT: Sofow, Rikki

TITLE OF INVENTION: MODIFIED ADENO ASSOCIATED VIRUS VECTOR

TITLE OF INVENTION: CAPABLE OF EXPRESSION FROM A NOVEL PROMOTER

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG
 STREET: 127 Peachtree Street, Suite 1200
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.2A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: 08/09,455,552A

FILING DATE: 31 May 1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: PeTTYman, David G.

REGISTRATION NUMBER: 63,448

REFERENCE/DOCKET NUMBER: 20004 0192

TELEPHONE: (404) 688 0770

TELEFAX: (404) 688 9880

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 1480 amino acids

TYPE: amino acid

STRANDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-455-552A-14

Query Match 5.48; Score 86; DB 2; Length 1480;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICSGSVKQVI-----GKHPHARVVCARKAKQIPEPEELSVKPKVNRNTHLNPQE 63

DB 220 CASAFCCGFLVIALFQA:G:GMMMKYINQAGKISER LVITSEMENIQSVKAYCW 277

QY 64 RQAFYRLLENEQTEFLSMDSCLRSISKYLLIAMVLAYFKRAAGLYI 109

DB 278 EAMKMKIENLRQTE-----LKLTRK---AAYVYFNSSAFFSGFEVVLVLPYAL 327

QY 110 -----SEYITNNFVVALYLANIMEDEEDYKYIEFWALGDSWRFQDFPIRPPFW 162

DB 328 IKGILKRIPTTISFQVILRMATRO-----FWAVQTWYDSLGAIKNTQ-DFL 375

QY 163 AKMYRAVVSRRCCDVMSKDP 185

DB 376 QKQYKILEYNIETTVVMEVNI 398

RESULT 14

US-07-890-609-2

Sequence 2, Application US/07890609

Patent No. 6001588

GENERAL INFORMATION:

APPLICANT: Tsui, Lap-Chew

APPLICANT: Rommings, Johanna M.

APPLICANT: Keren, Bat Sheva

TITLE OF INVENTION: Infrons and Exons of the Cystic Fibrosis Gene and

TITLE OF INVENTION: Mutations at Various Positions of the Gene

FILE REFERENCE: 3477-61, 044777/19840

CURRENT APPLICATION NUMBER: US/07/890,609

CURRENT FILING DATE: 1992-07-14

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO: 2

LENGTH: 1480

TYPE: PRT

ORGANISM: Homo sapiens

US-07-890-609 2

Query Match: 5.4%; Score: 66; IP: 3; Length: 1480;
Best Local Similarity: 20.7%; Field: 5; 2
Matches: 42; Conservative: 37; Mismatches: 74; Indels: 50; Gaps: 8;
QY 5 RASGSGSVKQV-----GAGHPPKRVVGGKAGQPPKPELISVAKKVKNTHNLQIQOE 63
DB 220 QASAPGQIGPLIVALEGAPGQMMKKYKPAK SEP--VITSEMIENIQSVKAYCW 277
QY 64 RQAFYPLENEQIQFELSMGSLPFDKYLAMVLZYKFAAGLYT----- 109
DB 278 EAMKMIENLRQF-----LKTFRK AAYVLYFNSSAFFSGFVVFVLSVLPYAL 327
QY 110 ---SYTNNFVALYANMDEEDYKYE FHWALGDSWKEELFPOFLERDDPW 162
DB 328 IKGILRKIFTTISIVLEMAVTE-----FWAV QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNYRNVSRRCCEVMSKDT 165
DB 376 OKQYKTLFNLTIFVVMNVT 168

RESULT 15
US-07-890-609-4
Sequence 4, Application: US/07890/ 9C
Patent No. 6001588
GENERAL INFORMATION:
APPLICANT: Tsui, Lap-chee
APPLICANT: Kommu, Jibanda M.
TITLE OF INVENTION: Mutations and variants of the Cystic Fibrosis Gene and
TITLE OF INVENTION: Mutations at various positions of the Gene
FILE REFERENCE: 3477-61, 0-4477, 4985
CURRENT APPLICATION NUMBER: US/07890/ 609C
CURRENT FILING DATE: 1992-07-13
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO: 4
LENGTH: 1480
TYPE: PRT
ORGANISM: Homo sapiens
US-07-890-609 4

Query Match: 5.4%; Score: 66; IP: 3; Length: 1480;
Best Local Similarity: 20.7%; Field: 5; 2
Matches: 42; Conservative: 37; Mismatches: 74; Indels: 50; Gaps: 8;
QY 5 RASGSGSVKQV-----GAGHPPKRVVGGKAGQPPKPELISVAKKVKNTHNLQIQOE 63
DB 220 QASAPGQIGPLIVALEGAPGQMMKKYKPAK SEP--VITSEMIENIQSVKAYCW 277
QY 64 RQAFYPLENEQIQFELSMGSLPFDKYLAMVLZYKFAAGLYT----- 109
DB 278 EAMKMIENLRQF-----LKTFRK AAYVLYFNSSAFFSGFVVFVLSVLPYAL 327
QY 110 ---SYTNNFVALYANMDEEDYKYE FHWALGDSWKEELFPOFLERDDPW 162
DB 328 IKGILRKIFTTISIVLEMAVTE-----FWAV QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNYRNVSRRCCEVMSKDT 165
DB 376 OKQYKTLFNLTIFVVMNVT 168

Search completed: July 31, 2002, 17:08:
Job time: 34.22 sec

The first part of the paper discusses the importance of the study and the objectives of the research. It highlights the need for a comprehensive understanding of the subject matter and the role of the researcher in this process. The second part of the paper presents the methodology used in the study, including the selection of participants, the data collection methods, and the analysis techniques. The third part of the paper discusses the results of the study and the conclusions drawn from the data. The final part of the paper provides a summary of the findings and discusses the implications of the study for future research.

Copyright (c) 1990-2000 Computer Aid.

OM protein protein search using BLAST

Run on: July 81 2002 15:06:00 - search time of 3 seconds
(without alignments)
543,148 million cell updates/sec

Title: US-09-889-592-2
Period score: 1632
Sequence: 1 MIMMUS-V79Amino VK2

Scoring table: BLOSUM62
Gapop 11.0, Gapext 0.5

Searched: 747574 seqs, 1167068 words
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 20000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 results

Database	Accession	Score	Length	Description
1	/S1155/peptide	16.6	434	An oocyte maturati
2	/S1155/peptide	16.6	434	An oocyte maturati
3	/S1155/peptide	16.6	434	An oocyte maturati
4	/S1155/peptide	16.6	434	An oocyte maturati
5	/S1155/peptide	16.6	434	An oocyte maturati
6	/S1155/peptide	16.6	434	An oocyte maturati
7	/S1155/peptide	16.6	434	An oocyte maturati
8	/S1155/peptide	16.6	434	An oocyte maturati
9	/S1155/peptide	16.6	434	An oocyte maturati
10	/S1155/peptide	16.6	434	An oocyte maturati
11	/S1155/peptide	16.6	434	An oocyte maturati
12	/S1155/peptide	16.6	434	An oocyte maturati
13	/S1155/peptide	16.6	434	An oocyte maturati
14	/S1155/peptide	16.6	434	An oocyte maturati
15	/S1155/peptide	16.6	434	An oocyte maturati
16	/S1155/peptide	16.6	434	An oocyte maturati
17	/S1155/peptide	16.6	434	An oocyte maturati
18	/S1155/peptide	16.6	434	An oocyte maturati
19	/S1155/peptide	16.6	434	An oocyte maturati
20	/S1155/peptide	16.6	434	An oocyte maturati
21	/S1155/peptide	16.6	434	An oocyte maturati
22	/S1155/peptide	16.6	434	An oocyte maturati

Pred. No. is the number of results credited by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Query Match	Score	Length	Description
1	99.6	16.6	434	An oocyte maturati
2	99.6	16.6	434	An oocyte maturati
3	99.6	16.6	434	An oocyte maturati
4	99.6	16.6	434	An oocyte maturati
5	99.6	16.6	434	An oocyte maturati
6	99.6	16.6	434	An oocyte maturati
7	99.6	16.6	434	An oocyte maturati
8	99.6	16.6	434	An oocyte maturati
9	99.6	16.6	434	An oocyte maturati
10	99.6	16.6	434	An oocyte maturati
11	99.6	16.6	434	An oocyte maturati

12	248.5	15.2	417	22	ABG02847	Novel human diapo
13	248.5	15.2	440	22	ABG21399	Novel human diapo
14	248.5	15.2	448	22	ABG14388	Novel human diapo
15	154.5	9.5	385	22	ABG14386	Novel human diapo
16	154.5	9.5	385	22	ABG16877	Novel human diapo
17	154.5	9.5	723	22	ABG0501	Novel human diapo
18	154.5	9.5	723	22	ABG02843	Novel human diapo
19	151	9.3	445	22	ABG0502	Novel human diapo
20	151	9.3	445	22	ABG02844	Novel human diapo
21	151	9.3	445	22	ABG14381	Novel human diapo
22	151	9.3	445	22	ABG16872	Novel human diapo
23	151	9.3	445	22	ABG21396	Novel human diapo
24	137	8.4	141	22	ABG05037	Human reproductive
25	135.5	8.3	1012	22	ABG02851	Novel human diapo
26	135.5	8.3	1012	22	ABG02851	Novel human diapo
27	135.5	8.3	1012	22	ABG16869	Novel human diapo
28	135.5	8.3	1012	22	ABG21403	Novel human diapo
29	127	7.8	900	22	ABG0511	Novel human diapo
30	127	7.8	1058	22	ABG02853	Novel human diapo
31	126	7.7	743	22	ABG21405	Novel human diapo
32	125	7.7	1619	22	ABG20707	Novel human diapo
33	99.5	6.1	501	22	ABG64959	Protophila melanog
34	98.5	6.0	485	22	ABG64960	Protophila melanog
35	93	5.7	81	22	ABH42024	Peptide #9529 euro
36	93	5.7	81	22	ABH25638	Protein #7637 euro
37	93	5.7	81	22	ABH63904	Human brain expres
38	93	5.7	81	22	ABH75716	Human bone marrow
39	93	5.7	81	22	ABH20672	Peptide #7106 euro
40	93	5.7	81	22	ABH35825	Peptide #9862 euro
41	93	5.7	177	22	ABG00497	Novel human diapo
42	93	5.7	177	22	ABG02840	Novel human diapo
43	93	5.7	236	22	ABG14379	Novel human diapo
44	93	5.7	236	22	ABG16871	Novel human diapo
45	93	5.7	463	22	ABH43636	Protophila melanog

ALIGNMENTS

RESULT	1	AA007752 standard. Protein; 300 AA.
AX	AA007752	
AX	AA007752	
AX	07 NOV 2000 (first entry)	
AX	At: oocyte maturation and proliferation protein Is26.	
AX	oocyte maturation, cell division, cell proliferation, cancer,	
AX	cell differentiation, cell differentiation, promoter, fertility treatment;	
AX	cell cycle progression.	
AX	Xenopus sp.	
AX	EF1026244-A1.	
AX	09-AUG-2000.	
AX	03-FEB-1999; 90EP-0102172.	
AX	03-FEB-1999; 90EP-0102172.	
AX	(EMBL-) EMBL EURO LAB MOLECULAR BIOLOGIE.	
AX	WPI: 2000-516092/47.	
AX	N PDB: AAA59327.	
AX	Polypeptide encoding a protein for inducing oocyte maturation and/or	
AX	promoting cell division, proliferation and differentiation useful for	
AX	treating cancer and disorders associated with fertility -	
AX	claim 6; Page 8; 17pp; English.	

DR WPI: 2001-639362/73.
 DR N-PSDB: AAS78522
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX (MYSE-) HVSFO INC.
 PS Claim 20: SEQ ID NO 44744; 103pp: English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010 AAG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 391 AA:
 SQ
 Query Match: 16.7%; Score 273; DE 22; Length 301;
 Best local Similarity 36.2%; Pred. No. 2, 1c 20;
 Matches 75; Conservative 32; Mismatches 72; Indels 28; Gaps 6;
 QY 45 PKAIPEREE-----SVKPMVFNTHLQAPDFQAFYPLENEQLIEFLSMGSLR 87
 DB 159 rklafepewkw hemisphkdkkkrtrssklypdeatflltdp/rlf/taakll: 218
 QY 88 ISDKYLIAWVLAYPEKRAAGLYISPTIMNFVALYANDMEDEPDYKYVIFPWALGDSW 147
 DB 219 vskylllam:laylar qqlfswqyqlhflalybandcedddedkq/itflfysk: 276
 QY 148 RELFPQLRLRDDFWAKMNYAVVS-----RNEDEFFFG 215
 DB 277 rsripllrkrirqlgysmnprraknrshiplvkrirrlqtrromprarkursjivifqkr 336
 QY 194 PMHH--SGAMRGYL---RNEDEFFFG 215
 DB 437 rrlfscmscravswspeeleentprq 363
 PS017 4
 ABG16876
 ID ABG16876 standard: protein: 367 AA
 XX
 AC ABG16876
 XX
 DT 18 FEB 2002 (first entry)
 XX
 DE Novel human diagnostic protein #168-7.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensics;
 KW food supplement; medical mapping; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W02017567-A2.
 XX

QY 11-OCF-2001.
 XX
 XX 30-MAR-2001; 2001WO-0508631.
 XX
 XX 21-MAR-2000; 2000US-0540317.
 XX
 XX 23-AUG-2000; 2000US-0949167.
 XX
 XX (MYSE-) HVSFO INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI: 2001-639362/73.
 XX N-PSDB: AAS81063.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 20: SEQ ID NO 47235; 103pp: English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010 AAG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 391 AA:
 SQ
 Query Match: 16.7%; Score 273; DE 22; Length 301;
 Best local Similarity 36.2%; Pred. No. 2, 1c 20;
 Matches 75; Conservative 32; Mismatches 72; Indels 28; Gaps 6;
 QY 45 PKAIPEREE-----SVKPMVFNTHLQAPDFQAFYPLENEQLIEFLSMGSLR 87
 DB 159 rklafepewkw hemisphkdkkkrtrssklypdeatflltdp/rlf/taakll: 218
 QY 88 ISDKYLIAWVLAYPEKRAAGLYISPTIMNFVALYANDMEDEPDYKYVIFPWALGDSW 147
 DB 219 vskylllam:laylar qqlfswqyqlhflalybandcedddedkq/itflfysk: 276
 QY 148 RELFPQLRLRDDFWAKMNYAVVS-----RNEDEFFFG 215
 DB 277 rsripllrkrirqlgysmnprraknrshiplvkrirrlqtrromprarkursjivifqkr 336
 QY 194 PMHH--SGAMRGYL---RNEDEFFFG 215
 DB 437 rrlfscmscravswspeeleentprq 363
 PS017 5
 ABG300498
 ID ABG300498 standard: protein: 367 AA.
 XX
 AC ABG300498
 XX

DI 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #489.
 DE Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX W0200175067-A2.
 PN 11-OCT-2001
 XX 30-MAR-2001; 2001WO-US08641.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSEQ) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI: 2001-639362/73.
 DR N-PSDB; AAS64685.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20: SEQ ID No 30857; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-AAG30377 represent novel human
 CC diagnostic amino acid sequences.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.
 XX Sequence 367 AA:

Query Match 16.6%; Score 271; DB 22; Length 367;
 Best Local Similarity 40.6%; Pred. No. 3,2e-20;
 Matches 71; Conservative 25; Mismatches 47; Indels 12; Gaps 4;
 QY 35 KKAQIDPEEEL-----SVKRWVNTNTHINIQPQKQAFYKLENEQIQEFLMSQSLR 87
 DE 160 RKLVAPEEELWAEMLQKMKLKRYSIVLPEHHEALNRLDPVKKRLAWDKDLR 219
 QY 88 ISDKYLAWVLAYPKRAAGLYTSEYTMNFFVALYANIMEDDELYKYETFPWALGDS- 146
 DE 228 VSDKYLAWVLAYFSR-GLPSWYQIRHFFALYLANDMEVDDEDPKQIYFYLYGKTR 278
 QY 147 -----WRELFPQ-----FLRLRDEFWAKNRYAVVSRRCDE 178
 DE 279 SRPLVNRTRFQICRLNPRKRSQIALFKIRFQFCSQGLAWVSREELCE 333

RES011 6
 AAG02841
 ID AAG02841 standard; Protein; 467 AA.
 XX AC AAG02841;
 XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #2842.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX W0200175067-A2.
 PN 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08641.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSEQ) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI: 2001-649462/74.
 DR N-PSDB; AAS67029.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20: SEQ ID No 34206; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detection or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-AAG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.
 XX Sequence 467 AA:

Query Match 16.6%; Score 271; DB 22; Length 467;
 Best Local Similarity 40.6%; Pred. No. 3,2e-20;
 Matches 71; Conservative 25; Mismatches 47; Indels 12; Gaps 4;
 QY 35 KKAQIDPEEEL-----SVKRWVNTNTHINIQPQKQAFYKLENEQIQEFLMSQSLR 87
 DE 160 RKLVAPEEELWAEMLQKMKLKRYSIVLPEHHEALNRLDPVKKRLAWDKDLR 219
 QY 88 ISDKYLAWVLAYPKRAAGLYTSEYTMNFFVALYANIMEDDELYKYETFPWALGDS 146

CC at ftp.wipo.int/pub/published_pet_sequences.

XX Sequence 434 AA:

Query Match 16.6% Score 271; DB 22; Length 444;

Best Local Similarity 40.6%; Pred. No. 40; 20;

Matches 71; Conservative 25; Mismatches 47; Gaps 4.

QY 35 RKQAIPERFET-----SVKPKVNVNTHLNLOPQRCATYETLENFCIGEFISMSGLR 87

DB 227 kvklapeeeiwaacmclgikmkikrrvslvlphecafnrlfcdpvtkrflawdkdlr 286

QY 88 ISDKYLLAMVLAVERKRAAGLYTSYTTMNIFFVALYLANDMEDDEYKYELFEWALGDS- 146

DB 287 vskkyllamvlaylar-advpswyqrhllfllalylandmceddab-qkquilylyqktr 345

QY 147 -----WRELEFPQ-----FLRLRDFWAKMNYRAVVSRCQDIE 178

DB 346 srplvrnrtrqrcrlnprarknsqialfqlrfccmsqrlawvsteelee 400

RESULT 9

ARG16875

10 ARG16875 standard; Protein; 434 AA.

XX ARG16875;

AC ARG16875;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #16866.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS W0200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08641.

PP 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSF-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YL;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS81062.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PP responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 47234; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. Amino acid sequences of the invention

CC diagnostic amino acid sequences of the invention.

CC Note: the sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from Wipo

CC at ftp.wipo.int/pub/published_pet_sequences.

XX Sequence 434 AA:

Query Match 16.6% Score 271; DB 22; Length 444;

Best Local Similarity 40.6%; Pred. No. 40; 20;

Matches 71; Conservative 25; Mismatches 47; Gaps 4.

QY 35 RKQAIPERFET-----SVKPKVNVNTHLNLOPQRCATYETLENFCIGEFISMSGLR 87

DB 227 kvklapeeeiwaacmclgikmkikrrvslvlphecafnrlfcdpvtkrflawdkdlr 286

QY 88 ISDKYLLAMVLAVERKRAAGLYTSYTTMNIFFVALYLANDMEDDEYKYELFEWALGDS- 146

DB 287 vskkyllamvlaylar-advpswyqrhllfllalylandmceddab-qkquilylyqktr 345

QY 147 -----WRELEFPQ-----FLRLRDFWAKMNYRAVVSRCQDIE 178

DB 346 srplvrnrtrqrcrlnprarknsqialfqlrfccmsqrlawvsteelee 400

RESULT 10

ARG16875

10 ARG16875 standard; Protein; 434 AA.

XX ARG16875;

AC ARG16875;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #16870.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS W0200175067-A2.

PN 11-OCT-2001.

PP 30-MAR-2001; 2001WO-US08641.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSF-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YL;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS81066.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 47234; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II). (I) is useful for treating disease states involving
 CC (II). (II) is useful for identifying antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC the polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AB00010-AB00037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.

XX Sequence 311 AA:

Query Match 15.7% Score 248.5; DB 22; Length 311.
 Best Local Similarity 44.7%; Pred. No. 9.8e-18;
 Matches 67; Conservative 21; Mismatches 45; Indels 17; Gaps 6.

QY 35 RKAOIPPEBEL-----SVKPKMWNTHNLQPQERQAFYRLLENQIQEFLSMUSCLR 87
 DB 54 RKLapepeelwvaemlcqlmkikrrvslvipelhealnrlldpvikrflawdkdir 113
 QY 88 ISDKYLIAVLAYFKRAAGLYTSEYTMNFFVALYLANDMEDEEDYKYEIFPWLALGDSW 147
 DB 114 VSDKYLIAVLAYFKRAAGLYTSEYTMNFFVALYLANDMEDEEDYKYEIFPWLALGDSW 171
 QY 148 RFLPFQFLRLKDDPW-----AMNPKAVVVSF 173
 DB 172 RSLPILIRK-----WQLQLRAMPRARKNR 197

RESULT 11

AB000505
 ID ARG000505 standard; Protein; 417 AA.

XX AC AB000505;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #496

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PP 31-MAR-2000; 2000US-0540217.

XX PP 23-AUG-2000; 2000US-0540117.

XX PA (HVSF) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX XX WPI: 2001-630362/7?

XX DR N-USDB; AAS64692.

XX PT New isolated polynucleotide and encoded polypeptides useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity.

XX Claim 20; SEQ ID NO 30864 101pp sequence

XX the invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity
 CC the polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AB00010-AB00037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.

XX Sequence 417 AA:

Query Match 15.2% Score 248.5; DB 22; Length 417;
 Best Local Similarity 29.3%; Pred. No. 9.8e-18;
 Matches 76; Conservative 32; Mismatches 78; Indels 74; Gaps 7;

QY 35 RKAOIPPEBEL-----SVKPKMWNTHNLQPQERQAFYRLLENQIQEFLSMUSCLR 87
 DB 160 RKLapepeelwvaemlcqlmkikrrvslvipelhealnrlldpvikrflawdkdir 219
 QY 88 ISDKYLIAVLAYFKRAAGLYTSEYTMNFFVALYLANDMEDEEDYKYEIFP----- 140
 DB 220 VSDKYLIAVLAYFKRAAGLYTSEYTMNFFVALYLANDMEDEEDYKYEIFP----- 278
 QY 141 -----WALGDSW-----RELFPQFLRLR 158
 DB 279 RSLPILIRK-----WQLQLRAMPRARKNR 338
 QY 159 RFLPFQFLRLKDDPW-----AMNPKAVVVSF 173
 DB 339 RFLPFQFLRLKDDPW-----AMNPKAVVVSF 173
 QY 212 RFLPFQFLRLKDDPW-----AMNPKAVVVSF 173
 DB 398 YSNANGR-----HQAG 408

RESULT 12

AB002847
 ID AB002847 standard; Protein; 417 AA.

XX AC AB002847;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #2838

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001

XX PP 30-MAR-2001; 2001WO-US08631.

XX PP 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 XX N-PSDB: AAS67034.
 DR
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 20: SEQ ID No 33206; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG3037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 417 AA;

Query Match 15.2%; Score 248.5; DB 2; Length 417;
 Best Local Similarity 29.3%; Pred. No. 9.8e-18;
 Matches 76; Conservative 32; Mismatches 78; Indels 73; Gaps 7;
 QY 35 RKAQIPEREL-----SVKPKVRNTHLNLOPQERQAFYRLLENEQTEFLSDSLR 87
 Db 160 rkvlapepecolwamolgikmkkrkrvstypethcafnrlledpqlk:fiawdkdlr 219
 QY 88 ISDKYLIAWVLAFFKKAAGLYTSRYTTMNFVVALYLANDMEDEDELYKYVEIPP----- 140
 Db 220 vskdyllamvlaytsr atpsswyqqrlnllalylandme-ebbskqqlthlylkykr 478
 QY 141 -----WALGDSW-----RELFPQFLR 158
 Db 279 srpllrkwtqlqshmprraknrslpllrkrfflyrstuprraknrslpllrkr 338
 QY 159 DDFWAKMNYRAVVSRRCCDEVMSKDPITW-----AWLRDRPMHHSAMRGYLENEDFP 211
 Db 339 fqlysmmsrraknrslq-vlfqkrffhfcsmcrswsspectecutqprdvdfqql 397
 QY 212 FDRPGHTPASCTLCHKAG 230
 Db 398 yshnqr-----bqaj 408
 RESULT 13
 ID ABG21399
 XX ABG21399 standard; Protein: 440 AA.
 AC
 XX ABG21399;
 XX
 DT 18-FEB-2002 (first entry)

XX
 DE Novel human diagnostic protein #21399.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensics;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 CS Homo sapiens.
 XX WO200175067 A2.
 PN 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08641.
 XX
 PR 31-MAR-2000; 2000US-0640217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 XX N-PSDB: AAS85586.
 XX
 CC New isolated polynucleotide and encoded polypeptides, useful in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits and to assess
 CC biodiversity
 XX
 PS Claim 20: SEQ ID No 51758; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30477 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 440 AA;

Query Match 15.2%; Score 248.5; DB 2; Length 440;
 Best Local Similarity 29.3%; Pred. No. 1.1e-17;
 Matches 76; Conservative 32; Mismatches 78; Indels 73; Gaps 7;
 QY 35 RKAQIPEREL-----SVKPKVRNTHLNLOPQERQAFYRLLENEQTEFLSDSLR 87
 Db 160 rkvlapepecolwamolgikmkkrkrvstypethcafnrlledpqlk:fiawdkdlr 219
 QY 88 ISDKYLIAWVLAFFKKAAGLYTSRYTTMNFVVALYLANDMEDEDELYKYVEIPP----- 140
 Db 220 vskdyllamvlaytsr atpsswyqqrlnllalylandme-ebbskqqlthlylkykr 478
 QY 141 -----WALGDSW-----RELFPQFLR 158
 Db 302 srpllrkwtqlqshmprraknrslpllrkrfflyrstuprraknrslpllrkr 361
 QY 159 DDFWAKMNYRAVVSRRCCDEVMSKDPITW-----AWLRDRPMHHSAMRGYLENEDFP 211

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG00377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.

XX
 SQ Sequence 385 AA:

Query Match 9.58; Score 154.5; DB 22; Length 385;
 Best Local Similarity 29.94; Pred. No. 1e-07;
 Matches 55; Conservative 21; Mismatches 61; Indels 47; Gaps 5;

QY 35 KKAQIPEREEL---SVKPKWVRNTHNLQPPHQAFYKLEENFQIQEFLSMDSCLR 87
 || || || | | | | | | | | | | | | | | | | | | | | | | | |
 Db 164 rkvlapoeeewvacmteqkkmkikrrrvslvlpchhcafurllcdpvrkrflawdkqir 222

QY 88 LSDKY-----LIAMVLAYEKRAAGLYISEYTTMNFVAL--YLANIMEELEEED 133
 :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 223 vsdkipsoptllqaspktilpasricirpsntrppsnrfpmstetpulsylandmcedded 282

QY 134 YKYELTFWALGDSWRKLELP-----QFLRLRDRDFWAKMNYR 168
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 283 pkqityilyakt--rsripllrkrrqleromprarkarsqivllqkrlqltesmscr 341

QY 169 AVVS 172
 | | |
 Db 342 awvs 345

Search completed: July 31, 2002, 17:07:23
 Job time: 5310 sec

electroporated into Top-10 F⁻, original library
 construction by Bruce Blumberg (Blumberg et al., 1991
 Science 253, 194-196; Hawlet et al., 1995, Genes Dev., 9,
 2934-2945). Normalized by Jihwan Song (Song, Cho and
 Blumberg, unpublished). Note: This is a Xenopus Gene
 Collection (XGC) library.*

```

BASE COUNT      138 a  158 c  159 g  137 t      1 others
ORIGIN

Query Match      45.0%; Score 551.2; DB 9; Length 593;
Best Local Similarity 97.9%; Pred. No. 8; 154;
Matches 569; Conservative 0; Mismatches 9; Indels 4; Gaps 1;

QY 116 tcttattatcaccatcttgaagatqatqatcctgaagcctgaagcctcctccttctt 175
Db 16 tccggtttatggcatttggacactgctggctaaagcgcctgaagagtggtgcttctt 75

QY 176 ctcaatcaacctcagccctggtatcccccttctacaaatgaagacatgaagatctaa 235
Db 76 ctcaatcaacctcagccctggtatcccccttctacaaatgaagacatgaagatctaa 135

QY 246 cccagagcagcctccatctatgacagcagggatgaagcagatcattgaagagcagcctc 295
Db 146 cccagagcagcctccatctatgacagcagggatgaagcagatcattgaagagcagcctc 195

QY 296 acgcccaggttcttgaagcagcagcagcagcagcagcagcagcagcagcagcagc 355
Db 196 acgcccaggttcttgaagcagcagcagcagcagcagcagcagcagcagcagcagc 255

QY 356 aacccaaatctgacaaatcaccatctcaatctacagccagcagcagcagcagcagc 415
Db 256 aacccaaatctgacaaatcaccatctcaatctacagccagcagcagcagcagcagc 315

QY 416 acagcctcctagaataatgaagcagcagcagcagcagcagcagcagcagcagcagc 475
Db 316 acagcctcctagaataatgaagcagcagcagcagcagcagcagcagcagcagcagc 375

QY 476 tttccagcaaatctctcaatgaagcagcagcagcagcagcagcagcagcagcagc 535
Db 476 tttccagcaaatctctcaatgaagcagcagcagcagcagcagcagcagcagcagc 432

QY 536 acacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 595
Db 433 acacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 492

QY 596 aggaagatgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 655
Db 493 aggaagatgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 552

QY 656 qtaagctttcccaaatcttttgcctcctcagcagcagcagcagcagcagcagcagc 696
Db 553 gtgagctttcccaaatcttttgcctcctcagcagcagcagcagcagcagcagcagc 593

RESULT 7
AL629647
LOCUS      AL629647 X97-gastrula Silurana tropicalis cDNA clone T84018120 57
DEFINITION mRNA sequence.
ACCESSION AL629647
VERSION    AL629647.1 GI:16599130
KEYWORDS   western clawed frog,
SOURCE     EST.
ORGANISM   Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Musobatrachia; Pipidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE  1 (bases 1 to 659)
            Buckle, E., Taylor, R., Ashurst, J., Zorn, A.M. and Rogers, J.
            Sanger Xenopus tropicalis EST project 2001 (10_2001)
            Unpublished (2001)
            Contact: Buckle, E.
  
```

Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE_ID: T84018120.sp6
 Sequencing primer: SP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.

FEATURES

```

Location/Qualifiers
Source      1.. 659
            /organism "Silurana tropicalis"
            /db_xref "taxon:8404"
            /clone "T84018120"
            /clone_lib "X97"
            /seq_start "gastrula (stages 30-33 (clawed))"
            /lab_host "Escherichia coli XL1 blue"
            /auto_vector "pcS107"
            /note "Site 1: EcoRI; Site 2: NotI; cDNA
            was cloned from 5' end of poly A+ RNA from stages
            10-13 gastrulae. EcoRI NotI and cDNA was then ligated
            into pcS107 with EcoRI at the 5' end and NotI at the 3'
            end."
BASE COUNT      168 a  165 c  159 g  167 t
ORIGIN

Query Match      43.6%; Score 528.6; DB 9; Length 619;
Best Local Similarity 96.4%; Pred. No. 6; 146;
Matches 588; Conservative 0; Mismatches 59; Indels 4; Gaps 2;

QY 131 tttaagcagatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 190
Db 13 ttgggcagctggcggctaaatggcctgaa cggcctggcctggcttcttctaaatggc 71

QY 191 qagccagctcctcctctctacatgaagcagcagcagcagcagcagcagcagcagc 250
Db 72 gatccggctggcggcctctctaaatgaagcagcagcagcagcagcagcagcagcagc 131

QY 251 tttatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 310
Db 132 ttttgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 191

QY 311 qagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 370
Db 192 cagcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 251

QY 371 qaaatcacatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca 430
Db 252 gcattacgcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca 311

QY 431 atagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 490
Db 312 atagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 371

QY 491 tttatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 550
Db 372 ttataacaaatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 429

QY 551 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 610
Db 429 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 488

QY 611 actatgaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca 670
Db 489 actatgaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca 548

QY 671 aatcttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 730
Db 549 aatcttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 608

QY 731 qagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 790
Db 609 gaaaatcttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 659
  
```



```

RESULT 8
RE490904
LOCUS B-19D12.x1 Placental/embryonic Xenopus laevis cDNA clone IMAGE340 (a) 3' similar to TR:Q9YGL1 Q9YGL1
DEFINITION SPEEDY PROTEIN. 1. mRNA sequence
ACCESSION RE490904
VERSION RE490904.1 GI:8613433
KEYWORDS EST
SOURCE African clawed frog
ORGANISM Xenopus laevis
REFERENCE 1 (bases 1 to 651)
AUTHORS Clifton,S., Johnson,S., Blumberg,P., Song,J., Hillier,L., Pape,D.,
        Martin,J., Wyllie,J., Underwood,K., Theisinger,H., Bowers,Y., Person
        ,R., Gibbons,M., Harris,N., Pitter,E., Jackson,Y., McGinnis,R.,
        Waterston,R. and Willson,R.
        WashU Xenopus EST project, 1999
        Unpublished (1999)
        Other_ESRs: db26c04.yl
        Contact: Sandy Clifton, Ph.D.
        WashU Xenopus EST project, 1999
        Washington University School of Medicine
        4444 Forest Park Parkway, Box 8561, St. Louis, MO 63108, USA
        Tel: 314 286 1800
        Fax: 314 286 1810
        Email: est@watson.wustl.edu
        Library constructed by and Sequenced by: Washington University
        Genome Sequencing Center
        Clone distribution: Xenopus clones from this library are available
        through the J.M.A.I.E. Consortium/ALN at: info@image.llnl.gov
        Seq primer: 400p from 3' end
        High quality sequence: yes
        Location/Qualifiers
            1..651
            Jordanref "Xenopus laevis"
            /def_xref "taxid:6115"
            /clone_image "SI001"
            /db_xref "Eukaryotic Genome Project"
            /library "WashU Xenopus EST project"
            /size "651bp"
            /size_type "nt"
            /cell_type "embryonic"
            /dev_stage "unfertilized egg"
            /alt_hist "Full length"
            /notes "Vector: pTZ19, Promoter: Site 2, NotI;
            PollA selected mRNA was prepared from unfertilized Xenopus
            laevis eggs. The library was constructed in the vector
            pTZ19-Pac as described by Ronaldo, M.F., Lennon, G. and
            Soares, M.B., Nucleic Acid Res. 17:111-126, 1989. Two
            approaches to ligation were used: one using a
            6721 bp, 19% GC, full length cDNA synthesis used a
            NotI-dBpI primer, double stranded cDNAs were ligated to
            EcoRI adapters digested with NotI, and directionally
            cloned into the NotI and PstI digested pTZ19-Pac vector.
            The library contained approximately 7.2 x 10^5
            recombinants, with average insert sizes of 1.5 kb."
            BASE COUNT 199 a 129 c 156 g 195 t 2 others
            ORIGIN
            Query Match: 32.6% Score 14.1 B 13. Length 651
            Best local Similarity 95.5% Ident 11-14
            Matches 595; Conservative 0; Mismatches 16; Indels 12; Gaps 6;
            QY 946 ctatcacagagatattttcttctacacagat ggaagatccagagagatctatgtttt 1005
            Db 651 ctatcacagagatattttcttctacacagat ggaagatccagagagatctatgtttt 593

```

clone distribution: Xenopus clone distribution information for this library can be found through research capabilities, visit their web page at: <http://www.reson.com/> please reference the id listed below when ordering this clone; Source lab clone id -

Seq primer: -400p from Gibco

High quality sequence stop: 411.

FEATURES

```

Location/Qualifiers
1..605
/organism="Xenopus laevis"
/zb_xref="taxon:8455"
/clone="XENOPUS_SOURCE_ID:"
/clone_lib="Xenopus laevis oocyte"
/tissue_type="oocyte (stages 5 and 6)"
/lab_host="Top-10 P"
/notes="Vector: phoscript SK+; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from 2nd of poly A+ RNA.
EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into Top-10 P+ original library
construction by Bruce Blumberg (Blumberg et al., 1991
Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
2923-2935). Normalized by Jihwan Song (Song, Cho and
Blumberg, unpublished). Note: this is a Xenopus Gene
Collection (XGC) library."
BASE COUNT 165 a 141 q 169 t
ORIGIN
Query Match 42.2%; Score 507.8; DB 9; Length 605;
Best Local Similarity 94.5%; Pred. No. 7.6e 140;
Matches 549; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

QY 739 tttataggttaatttccaaagatcccaactcattggagctggaatgaagatgagccatg 798
DB 1111111111111111111111111111111111111111111111111111111
DB 605 TCGTATGAGCATTTGTCGAAAGATCCACTACATTTGCTTGAAGAAATATGCGGCTC 546
QY 799 catcacacacacacacacacacacacacacacacacacacacacacacacacac 858
DB 1111111111111111111111111111111111111111111111111111111
DB 545 CATCACACGGGACCAACCGCGGTACTTTAGAAAACAAATATCTTTCCCGCGGT 486
QY 859 ccaggctttacacacacacacacacacacacacacacacacacacacacacacac 918
DB 1111111111111111111111111111111111111111111111111111111
DB 485 ACAGAGCTTACACACACACACACACACACACACACACACACACACACACACAC 426
QY 919 ggggttccccacacacacacacacacacacacacacacacacacacacacacac 978
DB 1111111111111111111111111111111111111111111111111111111
DB 425 GCGGTCGCCACACACACACACACACACACACACACACACACACACACACACAC 466
QY 979 ggggttccccacacacacacacacacacacacacacacacacacacacacacac 1038
DB 1111111111111111111111111111111111111111111111111111111
DB 365 GAGTCGCCACACACACACACACACACACACACACACACACACACACACACAC 306
QY 1039 catcacttacacacacacacacacacacacacacacacacacacacacacacac 1098
DB 1111111111111111111111111111111111111111111111111111111
DB 305 CATCAGCTTACACACACACACACACACACACACACACACACACACACACACAC 246
QY 1099 gaattgaacacacacacacacacacacacacacacacacacacacacacacac 1158
DB 1111111111111111111111111111111111111111111111111111111
DB 245 GAATGGGACACACACACACACACACACACACACACACACACACACACACACAC 186
QY 1159 aaactgaacacacacacacacacacacacacacacacacacacacacacacac 1218
DB 1111111111111111111111111111111111111111111111111111111
DB 185 AAGTCAGGACACACACACACACACACACACACACACACACACACACACACACAC 126
QY 1219 ggaagaatgaatattcc--- gaacatatatgaacaggaatattgaacaggaatatt 1274
DB 1111111111111111111111111111111111111111111111111111111
DB 125 GCGAAGTGAATATTCGACAGACAGAGATGATGAGCGGGAATAATGTCGACAGAAAGTGT 67

```

QY 1275 tttataggttaatttccaaagatcccaactcattggagctggaatgaagatgagccatg 1415

DB 111

DB 66 TTGAGTCGTCGAC 74

RESULT 10

AW782804/c

LOCUS

DEFINITION

AW782804

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AW782804

AW782804.1

EST

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.

1 (bases 1 to 568)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Papay, D., Martin, J., Wylie, T., Underwood, K., Theisinger, B., Bowers, Y., Pearson, J.B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McClure, R., Waterston, R., and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63106, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by Bruce Blumberg

Library normalized by Jihwan Song

DNA Sequencing by Washington University Genome Sequencing Center

Clone distribution: Xenopus clone distribution information for

this library can be found through research capabilities, visit their

web page at: <http://www.reson.com/> please reference the id listed

below when ordering this clone; Source lab clone id

Seq primer: -400p from Gibco

High quality sequence stop: 429.

Location/Qualifiers

1..568

/organism="Xenopus laevis"

/zb_xref="taxon:8455"

/clone="XENOPUS_SOURCE_ID:"

/clone_lib="Xenopus laevis oocyte"

/tissue_type="oocyte (stages 5 and 6)"

/lab_host="Top-10 P"

/notes="Vector: phoscript SK+; Site 1: EcoRI; Site 2:

XhoI; cDNA was prepared from 2nd of poly A+ RNA.

EcoRI-XhoI cut cDNA was then ligated into UniZap-XR

(Stratagene) with EcoRI at the 5' end and XhoI at the 3'

end. SS library phagemids were prepared by mass excision

from the original library and normalized by hybridization

to biotinylated driver (prepared from the same library by

PCR) to Cot-omega of 11. After removal of hybrids and

excess driver by streptavidin sepharose chromatography,

the ss-phagemids were made double stranded and

electroporated into Top-10 P+ original library

construction by Bruce Blumberg (Blumberg et al., 1991

Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,

2923-2935). Normalized by Jihwan Song (Song, Cho and

Blumberg, unpublished). Note: this is a Xenopus Gene

Collection (XGC) library."

153 a 130 c 156 q 149 t

ORIGIN

Query Match

Best Local Similarity

Score 501.8; DB 9; Length 568;

97.5%; Pred. No. 4.5e 130;

through the L.M.A.C.F. Bioscience/BLI at: info@bio.illinois.gov
Seq primer: 4981 from 3' to 5'
High quality sequence stop: 472

FEATURES

Source: 1: 552

Location/Qual: 111

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

through the L.M.A.C.F. Bioscience/BLI at: info@bio.illinois.gov
Seq primer: 4981 from 3' to 5'
High quality sequence stop: 472

FEATURES

Source: 1: 552

Location/Qual: 111

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

through the L.M.A.C.F. Bioscience/BLI at: info@bio.illinois.gov
Seq primer: 4981 from 3' to 5'
High quality sequence stop: 472

FEATURES

Source: 1: 552

Location/Qual: 111

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

through the L.M.A.C.F. Bioscience/BLI at: info@bio.illinois.gov
Seq primer: 4981 from 3' to 5'
High quality sequence stop: 472

FEATURES

Source: 1: 552

Location/Qual: 111

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

1: 552

```

Db 121 GATCTCGGGACGAGCTTCTGGGTAAATGAATACATACAGACAGTCTTATATGAAAGTG 180
QY 738 ctgtatgaatgaatgtccaaagatcccaactcattgaactgaactgaatgaatgaatgaat 797
Db 181 GGTGATGAGGAAAGTCTGAAAGATCCACATGCGCCCTGGGTGACAGATGCGCCGCA 240
QY 798 gcatcacagaagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 857
Db 241 CCAATCACAGCGAGGCAAGCGTGGGTACCTTACAAAGTACGAGTACGAGTACGAGTACGAG 300
QY 858 tccaaacttaccacagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 917
Db 301 TCGAGGCTTACAGCAGGAGCTGTACAGCTTGGCCATTAAGAGCAGGCTGTGTGAGTCTGG 360
QY 918 tgggtctcccaacaaactctctctcccaacaaacaaacaaacaaacaaacaaacaaacaaac 977
Db 361 TGGGCTCTCCACAAACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 978 aaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1012
Db 421 GAGTGGTGGAGGAGCTTCTGATGCTTGGCCGCTG 455

RESULT 15
B1444496 534 bp mRNA linear EST 21-AUG-2001
LOCUS B1444496
DEFINITION de25e07 y1 Wellcome CRC pRNA oocyte Xenopus laevis cDNA clone
IMAGE: 3472884 5' similar to IR:Q9VGL1 Q9VGL1 SPEEDY PROTEIN. ;
mRNA sequence.
ACCESSION B1444496
VERSION B1444496.1 GI:15269203
KEYWORDS EST
SOURCE African clawed frog.
ORGANISM
Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Pipidae;
Xenopodidae; Xenopus.
1 (bases 1 to 534)
Clifton,S., Johnson,S.L., Humberg,H., Song,J., Hillier,L., Pape,D.,
Martin,L., Wyke,T., Underwood,K., Theising,B., Powers,Y., Person,
J.B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterson,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 429.
FEATURES
Location/Qualifiers
1..534
/oranism "Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3472884"
/clone_lib="Wellcome CRC pRNA oocyte"
/trisun_type="oocyte"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Sequencing
according to Newkooop and Faber. Library was constructed
by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon
(Wellcome/CRC Institute)."
BASE COUNT 137 a 144 c 129 g 134 t 1 others
ORIGIN

```

```

Query Match 27.7% Score 4.06 E-06 DB 10; Length 544;
Best Local Similarity 90.8% Pred. No. 1.2e-118;
Matches 476; Conservative 0; Mismatches 45; Indels 4; Gaps 1;
QY 131 tttgagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 190
Db 14 tttgagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 190
QY 193 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 250
Db 74 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 250
QY 251 tttgagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 310
Db 134 tttgagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 310
QY 311 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 370
Db 194 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 370
QY 371 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 430
Db 254 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 430
QY 431 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 490
Db 314 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 490
QY 491 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 550
Db 374 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 550
QY 551 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 610
Db 431 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 610
QY 611 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 670
Db 451 gggagagagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 670

```

Search completed: July 31, 2002, 16:11:00
Job time: 2482 sec

Query Match
 Best Local Similarity 1.68; Score 37.2; DB 1; Length 7218;
 Matches 6; Conservative 208; Mismatches 156; Indels 0; Gaps 0;

QY 843 ctttttcccggttccagcttaccagccagctgtacacttgcacataaagagg 902
 Db 1075 YY 1134
 QY 903 tgtctgtacttctgtgggtctccacacacactctctctccacacagagattt 962
 Db 1135 YY 1194
 QY 964 tcaclacaccaataggagatgaatcccaagagactctcatalactcccttgaacttga 1022
 Db 1195 YY 1254
 QY 1023 quatcccgagttactgaacttaccacattctccagagccattgaattacagcc 1082
 Db 1255 YY 1314
 QY 1083 agatggagcgctgtgaatagccacacttgaatagccgattctctcccgagacttt 1142
 Db 1315 YY 1374
 QY 1143 attctctctactcacaagctcagacttattctctctctcctcctcagacttgcacat 1202
 Db 1375 YY 1434
 QY 1203 agacttaatt 1212
 Db 1435 GTAGCAAT 1444

RESULT 2
 US-08-998-416-224
 : Sequence 224, Application US/08598416
 : Patent No. 6239264
 : GENERAL INFORMATION:
 : APPLICANT: Philipson, Peter
 : APPLICANT: Pohlmann, Rainer
 : APPLICANT: Steibner, Sabine
 : APPLICANT: Mohr, Christine
 : APPLICANT: Wendland, Jürgen
 : APPLICANT: Knechtle, Philipp
 : APPLICANT: Reibschung, Corinne
 : TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
 : TITLE OF INVENTION: AND USES THEREOF
 : NUMBER OF SEQUENCES: 1152
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 6239264artis Corporation
 : STREET: 3054 Cornwallis Road
 : CITY: Research Triangle Park
 : STATE: No. 6239264th Carolina
 : COUNTRY: USA
 : ZIP: 27709
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/998,416
 : FILING DATE: 24-DEC-1997
 : CLASSIFICATION: 435
 : PRIORITY APPLICATION DATA:
 : APPLICATION NUMBER: 01/0016/97
 : FILING DATE: 31-DEC-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Mcigs, J. Timothy
 : REGISTRATION NUMBER: 38,241
 : REFERENCE/DOCKET NUMBER: PE/5-30306/A/CSC197h
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 919-541-8587

TELEFAX: 919-541-8684
 : INFORMATION FOR SEQ ID NO: 224:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 782 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : ORIGINAL SOURCE:
 : ORGANISM: PAC12080P
 : US-08-998-416-224

Query Match
 Best Local Similarity 59.5%; Score 34.8; DB 4; Length 782;
 Matches 112; Conservative 0; Mismatches 107; Indels 6; Gaps 1;

QY 1247 taacacttacttctgaataaataatataatctctctgaactgaactgaactgaact 1346
 Db 272 TAAATAGTATTATTAATTAAATTAATAATAATATATGATTAATAATAATATTA 341
 QY 1347 caatgaataacttctgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1409
 Db 332 TTAAGTTTAAATATTAATTAAGTATTAATAATAATAATAATAATAATAATAATA 408
 QY 1407 ctccacaaatgaacttactgaactgaacttctgaactgaacttctgaacttct 1466
 Db 489 CTTCCTTGATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 448
 QY 1467 qccacaaatattctgaataatgaacttctgaacttctgaacttctgaacttct 1508
 Db 445 AGTAATATTTTATTAAATATGATATATATATTAATTAATTAATTAATTAATTA 490

RESULT 3
 US-08-475-035-4
 : Sequence 3, Application US/08475035
 : Patent No. 5985554
 : GENERAL INFORMATION:
 : APPLICANT: KING, C. R.
 : APPLICANT: KRAUS, MATTHIAS H.
 : APPLICANT: ARMONSON, STUART A.
 : TITLE OF INVENTION: HUMAN GENE RELATED TO HOT INSTINCT PHOM
 : TITLE OF INVENTION: FGF RECEPTOR GENE
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 : STREET: Suite 1200, 127 Peachtree Street
 : CITY: Atlanta
 : STATE: Georgia
 : COUNTRY: USA
 : ZIP: 30304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: us/08/475,035
 : FILING DATE: 7 Jun 1995
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Portymann, David G.
 : REGISTRATION NUMBER: 33,438
 : REFERENCE/DOCKET NUMBER: 1414, 036
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 404/688-0770
 : TELEFAX: 404/688-9880
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 5542 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single

SEQUENCE CHARACTERISTICS:
 LENGTH: 665 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1581RF
 US-08-998-416-701

Query Match 2.09, Score 22.2, DB 4, Length 665;
 Best Local Similarity 50.2%, Pred. No. 2.4;
 Matches 108; Conservative 0; Mismatches 103; Indels 4; Gaps 1;
 QY 1287 aaacacgttgcttgaataaataatataatgtctgtgtggtgtaataaagat 1346
 DB 272 TAATAGTATTAAATTTAAATTAATAATAATTTTAAATTAATAATAATA 331
 QY 1347 caatgaacacgttgcaatgaatgaatgaatgaatgaatgaatgaatgaat 1406
 DB 332 TTAAGTTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATA 391
 QY 1407 ctccac 1466
 DB 392 CATTATATATATAA----TTATTAATGATGATTTTAAATTAATAATAATA 447
 QY 1467 gcac 1501
 DB 448 AGTAATATTTCTTATTTAAATAGTCTACCCCTTAAAT 482

RESULT 14
 US-08-998-416-701
 Sequence 701, Application US-08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jurgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Reibischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESS: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0 Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-08/998-416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: 14/5-30306/A/C001976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8609

INFORMATION FOR SEQ ID NO: 701:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 701 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1462RP
 US-08-998-416-701

Query Match 2.09, Score 22.2, DB 4, Length 701;
 Best Local Similarity 50.2%, Pred. No. 2.4;
 Matches 108; Conservative 0; Mismatches 103; Indels 4; Gaps 1;
 QY 1287 aaacacgttgcttgaataaataatataatgtctgtgtggtgtaataaagat 1346
 DB 272 TAATAGTATTAAATTTAAATTAATAATAATTTTAAATTAATAATAATAATA 331
 QY 1347 caatgaacacgttgcaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1406
 DB 332 TTAAGTTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATA 391
 QY 1407 ctccac 1466
 DB 392 CATTATATATATAA----TTATTAATGATGATTTTAAATTAATAATAATA 447
 QY 1467 gcac 1501
 DB 448 AGTAATATTTCTTATTTAAATAGTCTACCCCTTAAAT 482

RESULT 14
 US-08-998-416-786
 Sequence 786, Application US-08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jurgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Reibischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESS: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0 Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-08/998-416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: 14/5-30306/A/C001976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587

DR P PSDB: AAB07752.

XX polynucleotide encoding a protein for inducing oocyte maturation and/or
 promoting cell division, proliferation and differentiation useful for
 treating cancer and disorders associated with fertility.

PS Claim 1: Page 6; 17pp; English.

XX the present sequence encodes a protein which induces oocyte maturation
 and promotes cell division, proliferation and differentiation. The
 protein is a cell differentiation promoter. Pharmacological compositions
 comprising the protein or polynucleotide are useful for treating cancer
 or other pathological situations with uncontrolled cell growth by
 promoting cell division, differentiation and proliferation and is also
 useful for fertility treatments by promoting oocyte maturation. The
 polynucleotides are useful as diagnostic markers for cell proliferation
 and differentiation and for hybridisation experiments to determine the
 amount of homologous nucleic acid sequences. They are also useful as a
 target for identification of drugs that block cell cycle progression,
 cell proliferation and differentiation.

XX Sequence 1575 BP; 406 A; 368 C; 374 G; 427 T; 0 other;

Query Match 100.0%; Score 1575; DB 21; Length 1575;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ttaacaggacttgcagctccagttgaagtttttttcaagagctccgcccgaatgctat 60
DB 1 ttaacaggacttgcagctccagttgaagtttttttcaagagctccgcccgaatgctat 60
QY 61 tttttttatttcagagagctataagagagcagacagaaagaaataggcgagttcttg 120
DB 61 tttttttatttcagagagctataagagagcagacagaaagaaataggcgagttcttg 120
QY 121 tttatgcccatctgaacaaatgagaaacccctaaagatgcttcttcttctcag 180
DB 121 tttatgcccatctgaacaaatgagaaacccctaaagatgcttcttcttctcag 180
QY 181 atcaacctcagcccgatctccctttctcaatgagacaaatgaatgataacccag 240
DB 181 atcaacctcagcccgatctccctttctcaatgagacaaatgaatgataacccag 240
QY 241 gccagctccattttgcagcggggtgaagcaggctcattggcagagagatccacacgc 300
DB 241 gccagctccattttgcagcggggtgaagcaggctcattggcagagagatccacacgc 300
QY 301 cgaatttttgcagcggggtgaagcaggctcattggcagagagatccacacgc 360
DB 301 cgaatttttgcagcggggtgaagcaggctcattggcagagagatccacacgc 360
QY 361 aaaaatgagcgaatcccaatctcaatctacagcccccagagagagagagagagagag 420
DB 361 aaaaatgagcgaatcccaatctcaatctacagcccccagagagagagagagagagag 420
QY 421 ctcctagaataatgagcagatcaggaaattcttctatgagctctgtctcaggaatttc 480
DB 421 ctcctagaataatgagcagatcaggaaattcttctatgagctctgtctcaggaatttc 480
QY 481 gacaaatgagcgaatcccaatctcaatctacagcccccagagagagagagagagagag 540
DB 481 gacaaatgagcgaatcccaatctcaatctacagcccccagagagagagagagagagag 540
QY 541 agcagatcacacacacatgaattcttcttcttcttcttcttcttcttcttcttctt 600
DB 541 agcagatcacacacacatgaattcttcttcttcttcttcttcttcttcttcttctt 600
QY 601 gatgaagaagactataaatataaatattcttcttcttcttcttcttcttcttcttctt 660
DB 601 gatgaagaagactataaatataaatattcttcttcttcttcttcttcttcttcttctt 660
QY 661 cttttcccaaatctttcagctcccgagagagagagagagagagagagagagagagag 720

```

```

DB 661 cttttcccaaatcttttcaagctcccgagagagagagagagagagagagagagagagag 120
QY 721 atttttaattcgaagatgctatgagagagagagagagagagagagagagagagagagag 180
DB 721 atttttaattcgaagatgctatgagagagagagagagagagagagagagagagagagag 180
QY 781 ctgaagatccacccatgcatcacacagagagagagagagagagagagagagagagagagag 240
DB 781 ctgaagatccacccatgcatcacagagagagagagagagagagagagagagagagagag 240
QY 841 caettttttttttttttttttttttttttttttttttttttttttttttttttttttttt 300
DB 841 caettttttttttttttttttttttttttttttttttttttttttttttttttttttttt 300
QY 901 ggttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 360
DB 901 ggttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 360
QY 961 ttttcaatcaccaatgagagagagagagagagagagagagagagagagagagagagag 420
DB 961 ttttcaatcaccaatgagagagagagagagagagagagagagagagagagagagagag 420
QY 1021 ctgaatcccaagatgagagagagagagagagagagagagagagagagagagagagagag 480
DB 1021 ctgaatcccaagatgagagagagagagagagagagagagagagagagagagagagag 480
QY 1081 ccaatgagagagagagagagagagagagagagagagagagagagagagagagagagagag 540
DB 1081 ccaatgagagagagagagagagagagagagagagagagagagagagagagagagagag 540
QY 1141 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 600
DB 1141 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 600
QY 1201 tcaagatcaatgagagagagagagagagagagagagagagagagagagagagagagag 660
DB 1201 tcaagatcaatgagagagagagagagagagagagagagagagagagagagagagagag 660
QY 1261 cccagagagagagagagagagagagagagagagagagagagagagagagagagagagag 720
DB 1261 cccagagagagagagagagagagagagagagagagagagagagagagagagagagagag 720
QY 1321 tctctatgctcagatcaatgagagagagagagagagagagagagagagagagagagag 780
DB 1321 tctctatgctcagatcaatgagagagagagagagagagagagagagagagagagagag 780
QY 1381 atgagatgagagagagagagagagagagagagagagagagagagagagagagagagag 840
DB 1381 atgagatgagagagagagagagagagagagagagagagagagagagagagagagagag 840
QY 1441 ttcttaatactacatttttttttttttttttttttttttttttttttttttttttttt 900
DB 1441 ttcttaatactacatttttttttttttttttttttttttttttttttttttttttttt 900
QY 1501 tttcaatgagagagagagagagagagagagagagagagagagagagagagagagagag 960
DB 1501 tttcaatgagagagagagagagagagagagagagagagagagagagagagagagagag 960
QY 1561 agatgagatgagagagagagagagagagagagagagagagagagagagagagagagag 1020
DB 1561 agatgagatgagagagagagagagagagagagagagagagagagagagagagagagag 1020

```

RESULT 2

AAA59428

ID AAA59428 Standard: cDNA; 1457 bp.

XX

AA59428;

XX 07-NOV-2000 (first entry)

XX cDNA encoding an oocyte maturation and proliferation protein (527).

XX
PN W0200175067-A2.
XX
XX 11 OCT-2001.
PD
XX
PF 30 MAR-2001; 2001WO-US08641.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
FA
XX
XX Dmanac RT, Liu C., Tang YT;
PI
XX
XX WPI: 2001-639462/73.
DR P-FSDB: ABG90505.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1: SEQ ID No 496; 10pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC the polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1322 BP; 313 A; 453 G; 392 G; 264 T; 0 other;

```

Query Match.      6.5%  Score 102; DB 23; Length 1322;
Best Local Similarity 66.3%  Prod. No. 5, He-21;
Matches 163; Conservative 0; Mismatches 80; Indels 3; Gaps 1;

QY 389 tacagccacgagcgcagcgcctctacaggctctctagaacatqacagatcaggaaat 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 tgcctcctgagacccacagagcctctacaggctctctcaggtctctctctctaaaat 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 449 tctctctatgaactctctctaaagattctccacaaatctccatagcaatqctctaaq 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 tcttgacctggacaaagatctcagggtctcggacagatctccctagctatggctaaq 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 509 catattttaagcgcagcgcgcctctacaccagcagctacacaaatgaattctcttq 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 760 catattttaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 569 ttgcctctatcttgcctaatgacatqagagaatgaagaagatataaatatgaaatct 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 tggctctctactcggcaatgacatgagagagacacagagatcccaacaaaacatct 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 629 tccctt 634
    ||| |||
Db 877 tccact 882
    ||| |||

```

XX	AA567044	standard; cDNA; 1322 bp.
XX	AA567044	
XX	14-FEB-2002	(first entry)
XX	DNA encoding novel human diagnostic protein #2608.	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder; SS.	
XX	Homo sapiens.	
XX	W0200175067-A2.	
XX	11 OCT-2001.	
XX	30 MAR-2001; 2001WO 05086-41	
XX	31 MAR-2000; 2000HS 0540217.	
XX	23 AUG-2000; 2000US 0644167.	
XX	(HUSE) HUSPO INC.	
XX	Brmanac RI, Liu C, Tang YT;	
XX	WO/ 2001-63946-273.	
XX	P FSOH; ARG02H47.	
XX	New isolated polynucleotide and encoded polypeptides, useful in	
XX	diagnostics, forensics, gene mapping, identification of mutations	
XX	responsible for genetic disorders or other traits and to assess	
XX	biodiversity	
XX	Claim 1; SEQ ID NO 2438; 10pp; English.	
XX	The invention relates to isolated polynucleotide (I) and	
XX	polypeptide (II) sequences. (I) is useful as hybridization probes,	
XX	polymerase chain reaction (PCR) primers, oligomers, and for chromo-	
XX	somal gene mapping, and in recombinant production of (II). The	
XX	polynucleotides are also used in diagnostics as expressed sequence	
XX	for identifying expressed genes. (I) is useful in gene therapy tech-	
XX	to restore normal activity of (II) or to treat disease states invol-	
XX	ving (II). (II) is useful for generating antibodies against it, detect-	
XX	quantifying a polypeptide in tissue, as molecular weight markers in	
XX	a food supplement. (II) and its binding partners are useful in medi-	
XX	cating of sites expressing (II). (I) and (II) are useful in medi-	
XX	c disorders involving aberrant protein expression or biological activ-	
XX	The polypeptide and polynucleotide sequences have applications in	
XX	diagnostics, forensics, gene mapping, identification of mutations	
XX	responsible for genetic disorders or other traits to assess biodive-	
XX	sity and to produce other types of data and products dependent on DNA	
XX	and amino acid sequences. AA564197 AAS94564 represent novel human	
XX	diagnostic coding sequences of the invention.	
XX	Note: The sequence data for this patent did not appear in the prior	
XX	art. Specification, but was obtained in electronic format directly from	
XX	at http://wipo.int/pub/published/pet_sequences .	
XX	Sequence 1322 BP; 313 A; 453 C; 192 G; 264 T; 0 other	

[illegible]

XX Homo sapiens.
OS
XX WO2007/007-A2.
XX
XX
XX
PD IWOCT 20 1.
XX
XX
XX 09 MAR 2007; 2001WO 05086-1.
XX
XX 31 MAR 2007; 2006US 054027.
XX
PR 23 AUG 2007; 2006US 06491-7.
XX
XX (HVS-1) HVSQ INC.
XX
XX formant F2, L1a C, L1aa Y1;
PL
PL WPI: 2007_639362/74.
DR DR P-FSDB: AAG02841.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity .
XX
XX Claim 1: SEQ ID NO 2832; 103pp; English
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligonucleotides, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used as forensic tools as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (ii) or to treat disease states involving
CC (ii). (II) is useful for generating antibodies against it, detecting or
CC quantifying a polypeptide in tissues as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging at sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of reagents dependent on DNA and
CC amino acid sequences. AAG4147 AAG51564 represent novel human
CC diagnostic coding sequences of the invention.
XX Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at <http://wipo.int/patpublic/txtseq.pl?seqno=639362>
XX Sequence :104 BP; 264 AA; 82 CT; 240 T; 204 T; 6 G; 6 C; 6 G; 6 C;

Query Match:	6.99	Score:	10.4	DB:	23	Length:	1104
Best Local Similarity:	66.00	Match:	5.2	Score:	20		
Motif:	16	Conserved Pos:	6	Nucleotide:	81	Tuples:	3
QY	389	tatataccacgaagcgcgccttatatatctcttataaatatagaqaatttaqaaat	448				
Db							
QY	572	tatatacttgagagcgagagagccttatataatggttttagagtccttatcatataaat	631				
QY	449	tactttctatcaaacctcttatataattttatcacatattatcctatggcaatctctaa	508				
Db							
QY	632	cctgaacctatgataatataatctgaatctttatataatattatctcttgaatgat	691				
QY	529	tatatattttaagggttatagagccttatataatattatgacacatgaattctctttg	568				
Db							
QY	692	tgattattcaagccgaacagagagccttgccttatgaataaccagaaatctattcttcc	748				
QY	559	ttaaataatattctcttatataatgaatgagagcttatatttatattatataatataat	628				
Db							
QY	749	tgaatcttccaacaatataatataagataagagcgaagagacccaacaataatctt	808				
QY	629	tcctat	634				

E 809 Contact 814

ASJIT 9
AAS85577 standard; cDNA; IL04 BP.

AAS85577:
13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #21381.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
Homo sapiens.

WZ000175067-A2.
11-OCT-2001.

30-MAR-2001; 2001WO-US098631.
31-MAR-2000; 2000GS-0548217.
23-AUG-2000; 2000GS-0649167.
(HYSP-) HYSKO INC.

Brahmac RT, Liu G, Tanq YT;
WEI, 2001-6/306273.
P-PSON; ABC21390.

New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -

Claim 1; SEQ ID NO 21381; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for thomosis
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence
for identifying expressed genes. (I) is useful in gene therapy tech-
to restore normal activity of (II) or to treat disease states involv-
(II). (II) is useful for generating antibodies against it; detecting
quantitating a polypeptide in tissue, as molecular weight markers in
a food supplement. (II) and its binding partners are useful in medi-
imaging of sites expressing (II). (I) and (II) are useful for treat-
disorders involving aberrant protein expression or biological activ-
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiver-
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS94564 represent novel human
diagnostic coding sequences. AAS94564 represent novel human
Note: The sequence data for this patent did not appear in the print
specification, but was obtained in electronic format directly from
at http://wipo.int/pub/published_pat_sequences.

Sequence 1104 has 264 A's, 252 C's, 349 G's, 219 T's, 0 other:

Query Match 6.4%; Score 100.4; DB 23; Length 1114;
Best local similarity 65.9%. Frac No. 1.6e-20;
Matches 162; Conservative 0; Mismatches 81; Indels 3; Gap

489 tacagcccaaggcgcagcgacctctcacagcgccttcagaaatgagaattcaggat 4
572 tactcccttaaacgtaccagcactttagacgaagtacttgagactcvtalalaaagat 11


```

FEATURES             location/Qualifiers
  source              1..1575
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    /clone="1"
    gene              214..1116
    CDS                214..1116
    /gene="ls26"
    /function="triggers oocyte meiotic maturation"
    /codon_start=1
    /product="p33 ringo"
    /protein_id="CA84295.1"
    /db_xref="GI:4995925"
    /translation="MRIMOSVTRASSICGSKVQVIGKHPIHARVVGARKAGIPEREE
LSVKPMRNTIILNPOEROAFYRLIENEOIQFLSMISCLIPISKYLIMVLAYPEK
RAAGLYSEYVTLNFEVALYLANDMEDEEDYKYLEFWALGDSWRELFPOFLKLRDD
FWAKMYRAVWSRRCDVMSKDPITHAWLRIKPMHHSCKAMKGYLRNDDPPRPQGL
TPASCTLHKGACVSGGVSNNSSPQETFIHYTNREWSQHLMLPELLLDPECTH
DRIHQPLVIGUPDGTALFWHHL"
BASE COUNT          406 a   368 c   374 g   427 t
ORIGIN

Query Match      100.0%; Score 1575; DB 5; Length 1575;
Best local Similarity 100.0%; Prod. No. 0;
Matches 1575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttaacaggacttccagctccatqtaagcttttttcaagaagctcccccacatctat 60
DB 1 ttaacaggacttccagctccagctccagctccagctccagctccagctccagct 60
QY 61 ttttttatttattccaggaagctataaagaagagagagagagagagagagagag 120
DB 61 ttttttatttattccaggaagctataaagaagagagagagagagagagagagag 120
QY 121 ttatcccatcttgcagctgaagctgaagctgaagctgaagctgaagctgaagct 180
DB 121 ttatcccatcttgcagctgaagctgaagctgaagctgaagctgaagctgaagct 180
QY 181 atcaacctccaggaagctgaagctgaagctgaagctgaagctgaagctgaagct 240
DB 181 atcaacctccaggaagctgaagctgaagctgaagctgaagctgaagctgaagct 240
QY 241 qccagctccatcttgcaggaaggaaggaaggaaggaaggaaggaaggaaggaag 300
DB 241 qccagctccatcttgcaggaaggaaggaaggaaggaaggaaggaaggaaggaag 300
QY 301 cgggtttgttgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
DB 301 cgggtttgttgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
QY 361 aaaaatqtlqgaagatgccatgccatgccatgccatgccatgccatgccatgccat 420
DB 361 aaaaatqtlqgaagatgccatgccatgccatgccatgccatgccatgccatgccat 420
QY 421 ctctcaaaaaaaagagatgaagaaatctctctctctctctctctctctctctct 480
DB 421 ctctcaaaaaaaagagatgaagaaatctctctctctctctctctctctctctct 480
QY 481 gacaagtatctcagatgaagatgttctagcatattttaaagcggcgcgcgcgcgc 540
DB 481 gacaagtatctcagatgaagatgttctagcatattttaaagcggcgcgcgcgcgc 540
QY 541 aqqaatcacacacacatgaattcttcttgccttgccttgccttgccttgccttgc 600
DB 541 aqqaatcacacacacatgaattcttcttgccttgccttgccttgccttgccttgc 600
QY 601 qatqaagaagactataaataaaatcttccctggagcactgaagaaatcttgccttgc 660
DB 601 qatqaagaagactataaataaaatcttccctggagcactgaagaaatcttgccttgc 660
QY 661 cttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720

```

```

DB 661 cttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
QY 721 qttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
DB 721 cttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
QY 781 ctgaagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
DB 781 ctgaagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
QY 841 gatttttcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
DB 841 gatttttcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
QY 901 qttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
DB 901 cgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
QY 961 tttcccatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
DB 961 tttcccatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
QY 1021 ctgaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
DB 1021 ctgaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
QY 1081 cccaaatgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
DB 1081 cccaaatgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
QY 1141 tttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1200
DB 1141 tttttcccaaatcttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1200
QY 1201 tccagcttaattgaagatgaagatgaagatgaagatgaagatgaagatgaagat 1260
DB 1201 tccagcttaattgaagatgaagatgaagatgaagatgaagatgaagatgaagat 1260
QY 1261 cccagagaagatgtttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320
DB 1261 cccagagaagatgtttttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320
QY 1321 tctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1380
DB 1321 tctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1380
QY 1381 atgtatgtgaagatctttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440
DB 1381 atgtatgtgaagatctttgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440
QY 1441 ttgcttaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1500
DB 1441 ttgcttaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1500
QY 1501 ttccaatgaagatgaagatgaagatgaagatgaagatgaagatgaagatgaagat 1560
DB 1501 ttccaatgaagatgaagatgaagatgaagatgaagatgaagatgaagatgaagat 1560
QY 1561 aaatgaagatgaagatgaagatgaagatgaagatgaagatgaagatgaagat 1575
DB 1561 aaatgaagatgaagatgaagatgaagatgaagatgaagatgaagatgaagat 1575

```

```

RESULT 2
AX033536 LOCUS
DEFINITION Sequence 1 from Patent EP1626244.
ACCESSION AX033536
VERSION AX033536.1 GI:16280282
KEYWORDS :
SOURCE :
ORGANISM Xenopus sp.

```


VERSION AJ249978.1 GI:6066823
 KEYWORDS Is26 gene; p33 rindq.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 1376)
 Ferby, I., Blazquez, M., Palmer, A., Eritja, R. and Nebreda, A.P.
 A novel p34(cdc2)-binding and activating protein that is necessary
 and sufficient to trigger G2/M progression in Xenopus oocytes
 Genes Dev. 13 (16), 2177-2189 (1999)
 99496721
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 1. 1376
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 214. .1113
 /gene="Is26/clone 5"
 214. .1113
 /gene="Is26/clone 5"
 /function="triggers oocyte meiotic maturation"
 /codon_start=1
 /product="p33 rindq"
 /protein_id="CAH58366.1"
 /db_xref="GI:6066823"
 /translation="MKHMSVIRALISGSGVKQVLSKQHPHNVGVGAKKAOIPEREE
 LSKPKMVENTILNLPQERQAFYLLFNELQKFI SMDSCTPISTLYLTMVLYEP
 RAGLYTSEITNMFVALYLANDMEDEEDKYEFPMALGDSWRELFQPLRLRDF
 WAKMNRVAVSRGCDVNSKPTTHWMTPEPTIDHSAMPVYIPREDSYSPFSLT
 PASCTLCHGACVDSGVSNNSSPFOEIPHYIKRWSQRLMLSPHLLIDPDCIHD
 LRIHQHPLVGLPEDCIHALEWHHL"
 BASE COUNT 358 a 331 c 336 g 351 t
 ORIGIN

Query Match 82.0%; Score 1291.2; DB %; Length 1376;
 Best local similarity 98.7%; Pred. No. 0;
 Matches 1346; Conservative 0; Mismatches 8; Indels 10; Gaps 4;

QY 1 tlaacagaaactlcaagelccagatlaagatlttlttcaagaagccagcccaatgcatat 60
 Db 1 TTAACAGGACTTCCAGCTCCAGCTCAGCTGAGCTTTTTCAGAGGCGGACCCCAATGCTAT 60
 QY 61 ttttttattattccagagagatataaagagagagagagagagagagagagagagagag 120
 Db 61 TTTTATTATTCCAGAGAGAGATATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 121 ttatagccatttggccattggttgcctgaagccctgaaggtgcatggttttttgcctg 180
 Db 121 TTTATGCCATTTTGGCCATTGGTGGCTGAAGCGCTTCAAGCTGCTGCTTCTTCTCCAG 180
 QY 181 atcaacccctggagccagatgctcccttcttcaaalgaagagagagagagagagagag 240
 Db 181 ATCAACCCCTGGGCGGCTGCTCCCTCTTCTACAAATGAGCCACATGACAGGTAAACCCG 240
 QY 241 gccagctccatttggag 300
 Db 241 GCCAGCTCCATTGTGGAG 300
 QY 301 cggagttgttggag 360
 Db 301 CGGAGTTGTGGAG 360
 QY 361 aaatagtgag 420
 Db 361 AAAATGCTCCGAAATACCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 420

QY 421 cctctcaagaatagaagagatctctctctctctctctctctctctctctctctctctct 480
 Db 421 CTCCTAGAAATAGAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
 QY 481 gacaaatct 540
 Db 481 GACAACTATCTCAATAGCAATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
 QY 541 agcag 600
 Db 541 AGCAG 600
 QY 601 gatgaagaagaatct 660
 Db 601 GATGAGAGAGAGATATAATATAGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
 QY 661 ctlttcccaagaatlttct 720
 Db 661 CTTTCCCAAGATTTCT 720
 QY 721 qtttct 780
 Db 721 GTTGTAGTGGAGAGTCT 780
 QY 781 ctgagagatcccccctgcatcagagagagagagagagagagagagagagagagagag 840
 Db 781 CTGAGAGATAGCTGCT 840
 QY 841 gacttttct 900
 Db 841 GACTTTCT 900
 QY 901 ggttct 960
 Db 901 GGTCT 960
 QY 961 tttct 1020
 Db 961 TTTCT 1020
 QY 1021 ctgaatcccaatgcatcag 1080
 Db 1021 CTGAACTCCAGAGTCT 1080
 QY 1081 ccagatgag 1140
 Db 1081 CCAGATGAG 1140
 QY 1141 ttattct 1200
 Db 1141 TTATCT 1200
 QY 1201 tcaaatct 1260
 Db 1201 TCAGACTTAATTCAGATGAG 1260
 QY 1257 tatgag 1316
 Db 1257 TATG GAG 1316
 QY 1316 -aatct 1376
 Db 1317 GAATGTTCT 1376

RESULT 4

XLA133500

Xenopus laevis

DEFINITION

Xenopus laevis mRNA

Accession

XLA133500.1

Version

1.27

Keywords

African clawed frog


```

Db 367 TGCTCCCTGACGACACACAGAGGCTTCAACAGGCTGTGAGGATATGTCATTAATAAGAT 426
449 Tcctttctatggaactctctctaaagatttcccaagaagtatctatagaatattctctaa 508
427 Tcctggccttggcacaacaaatcagacagggctgagcacaagatctctctctctctctct 511
509 cactatcttlaacacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacga 568
487 CGTATTTGACCGG---AGCGGGCTTCGCTCTGCGAATACCAACGCTTCATTTCTTCC 543
569 ttgcctctatctgctaatcactagaagaatgaagaagaactataatagaatct 628
544 TGCTCTCTACTCGGCAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 603
629 tccctt 634
604 TCCACT 609

RESULT 8
AC097724
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-713319, *** SEQUENCING IN
PROGRESS ***, 27 unordered pieces.
ACCESSION AC097724
VERSION AC097724.3 GI:18250175
KEYWORDS HIC; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 200975)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 200975)
Waterston,R.H.
Direct Submission
Submitted (22-Oct-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jan 19, 2002 this sequence version replaced gi:18158408.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/genseq/chr2.xshtml
----- Project Information -----
Center project name: H.NI0714D19
----- Summary Statistics -----
Sequencing vector: M13; 32x
Chemistry: Dye-primer ET; 0x of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 188781 bases at least Q40
Consensus quality: 190782 bases at least Q30
Consensus quality: 191907 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1 88: contig of 88 bp in length
89 188: gap of unknown length
189 1734: contig of 1545 bp in length
1734 1834: gap of unknown length
1834 4044: contig of 1201 bp in length
4044 4144: gap of unknown length
3035

```

```

FEATURES
Source
1..200975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-713319"
1..88
/feature="assembly_name:Contig1"
189..1734
/feature="assembly_name:Contig449"
1834..4034
/feature="assembly_name:Contig54"
4145..4565
/feature="assembly_name:Contig58"
4666..5914
/feature="assembly_name:Contig61"
6015..7560
/feature="assembly_name:Contig62"
7661..9687
/feature="assembly_name:Contig64"
9788..11410
/feature="assembly_name:Contig64"
11511..14329
/feature="assembly_name:Contig65"
14430..17747
/feature="assembly_name:Contig66"

```


